

Lecture 7

Last time: Moments of the Poisson distribution from its generating function.

$$G(s) = e^{\mu(s-1)}$$

$$\frac{dG}{ds} = \mu e^{\mu(s-1)}$$

$$\frac{d^2G}{ds^2} = \mu^2 e^{\mu(s-1)}$$

$$\bar{X} = \left. \frac{dG}{ds} \right|_{s=1} = \mu$$

$$\overline{X^2} = \left. \frac{d^2G}{ds^2} \right|_{s=1} + \left. \frac{dG}{ds} \right|_{s=1}$$

$$= \mu^2 + \mu$$

$$\sigma^2 = \overline{X^2} - \bar{X}^2$$

$$= \mu^2 + \mu - \mu^2$$

$$= \mu$$

$$= \bar{X}$$

Example: Using telescope to measure intensity of an object

Photon flux \rightarrow photoelectron flux. The number of photoelectrons are Poisson distributed. During an observation we cause N photoelectron emissions. \bar{N} is the measure of the signal.

$$S = \bar{N} = \lambda t$$

$$\sigma_N^2 = \mu = \lambda t$$

$$\frac{S}{\sigma_N} = \frac{\lambda t}{\sqrt{\lambda t}} = \sqrt{\lambda t}$$

$$t = \frac{1}{\lambda} \left(\frac{S}{\sigma_N} \right)^2$$

For signal-to-noise ratio of 10, require $\bar{N} = 100$ photoelectrons. All this follows from the property that the variance is equal to the mean.

This is an unbounded experiment, whereas the binomial distribution is for n number of trials.

3. The Poisson Approximation to the Binomial Distribution

The binomial distribution, like the Poisson, is that of a random variable taking only positive integral values. Since it involves factorials, the binomial distribution is not very convenient for numerical application.

We shall show under what conditions the Poisson expression serves as a good approximation to the binomial expression – and thus may be used for convenience.

$$b(k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

Consider a large number of trials, n , with small probability of success in each, p , such that the mean of the distribution, np , is of moderate magnitude.

Define $\mu \equiv np$ with n large and p small

$$p = \frac{\mu}{n}$$

Recalling:

$$n! \sim \sqrt{2\pi n} n^{n+\frac{1}{2}} e^{-n} \text{ Stirling's formula}$$

$$e^{-\mu} = \lim_{n \rightarrow \infty} \left(1 - \frac{\mu}{n}\right)^n$$

$$\begin{aligned} b(k) &= \frac{n!}{k!(n-k)!} \frac{\mu^k}{n^k} \left(1 - \frac{\mu}{n}\right)^{n-k} \\ &\approx \frac{\mu^k}{k!} \frac{\sqrt{2\pi n} n^{n+\frac{1}{2}} e^{-n}}{\sqrt{2\pi} (n-k)^{n+\frac{1}{2}} e^{-n+k} n^k} \left(1 - \frac{\mu}{n}\right)^{n-k} \\ &= \frac{\mu^k}{k!} \frac{n^{n+\frac{1}{2}} e^{-n}}{n^{n+\frac{1}{2}} \left(1 - \frac{k}{n}\right)^{n-k+\frac{1}{2}} e^k} \left(1 - \frac{\mu}{n}\right)^{n-k} \\ &\approx \frac{\mu^k}{k!} \frac{e^{-\mu}}{e^{-k} e^k} \text{ as } n \text{ becomes large relative to } k \\ &= \frac{1}{k!} \mu^k e^{-\mu} \end{aligned}$$

The relative error in this approximation is of order of magnitude

$$\text{Rel. Error} \sim \frac{(k - \mu)^2}{n}$$

However, for values of k much smaller or larger than μ , the probability becomes small.

The Normal Distribution

Outline:

1. Describe the common use of the normal distribution
2. The practical employment of the Central Limit Theorem
3. Relation to tabulated functions
 - Normal distribution function
 - Normal error function
 - Complementary error function

1. Describe the common use of the normal distribution

Normally distributed variables appear repeatedly in physical situations.

- Voltage across the plate of a vacuum tube
- Radar angle tracking noise
- Atmospheric gust velocity
- Wave height in the open sea

2. The practical employment of the Central Limit Theorem

X_i ($i = 1, 2, \dots, n$) are independent random variables.

Define the sum of these X_i as

$$S = \sum_{i=1}^n X_i$$

$$\bar{S} = \sum_{i=1}^n \bar{X}_i$$

$$\sigma_S^2 = \sum_{i=1}^n \sigma_{X_i}^2$$

Then under the condition

$$\lim_{n \rightarrow \infty} \frac{\beta}{\sigma_S^3} = 0$$

$$\beta = \sum_{i=1}^n \beta_{X_i}$$

$$\beta_{X_i} = \left| \overline{(X_i - \bar{X}_i)^3} \right|$$

the limiting distribution of S is the normal distribution. Note that this is true for any distributions of the X_i .

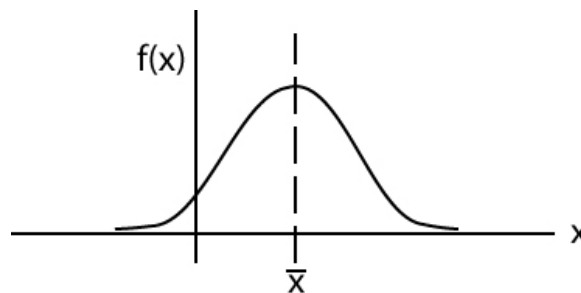
These are sufficient conditions under which the theorem can be proved. It is not clear that they are necessary.

Notice that each of the noises mentioned earlier depend on the accumulated effect of a great many small causes e.g., voltage across plate: electrons traveling from cathode to plate.

It is convenient to work with the characteristic function since we are dealing with the sum of independent variables.

Normal probability density function:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$



Normal probability distribution function:

$$\begin{aligned} F(x) &= \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(u-m)^2}{2\sigma^2}} du \\ &= \int_{-\infty}^{\frac{x-m}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} dv \end{aligned}$$

Where:

$$m = \bar{X}$$

$$v = \frac{u-m}{\sigma}$$

$$dv = \frac{1}{\sigma} du$$

This integral with the integrand normalized is tabulated. It is called the *normal probability function* and symbolized with Φ .

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} dv$$

This is a different x . Note the relationship between this and the quantity x previously defined. We use x again here as this is how Φ is usually written.

Not only this function but also its first several derivatives which appear in analytic work are tabulated.

3. Relation to tabulated functions

Even more generally available are the closely related functions:

Error function:

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du$$

Complementary error function:

$$\text{cerf}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-u^2} du$$

$$\Phi(x) = \frac{1}{2} \left[1 + \text{erf} \left(\frac{x}{\sqrt{2}} \right) \right]$$

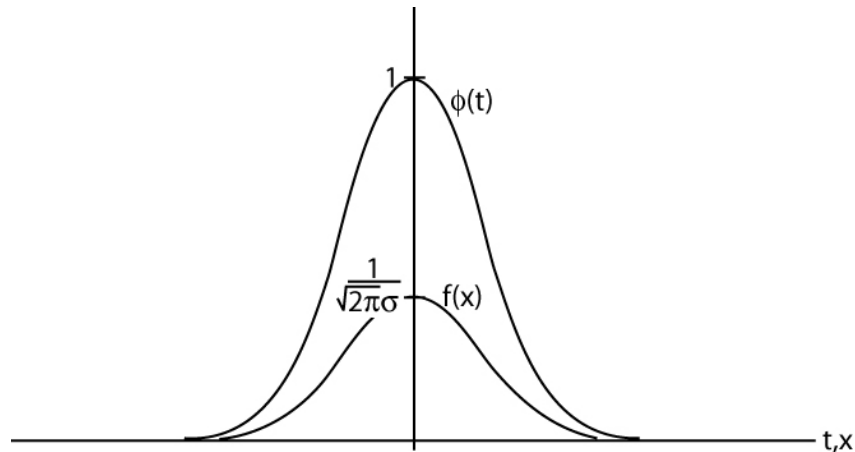
$$\begin{aligned} \phi(t) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{jtx} e^{-\frac{(x-m)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{jt(m+\sigma y)} e^{-\frac{y^2}{2}} dy, \text{ where } y = \frac{x-m}{\sigma} \\ &= \frac{1}{\sqrt{2\pi}} e^{jtm} \int_{-\infty}^{\infty} (\cos(t\sigma y) + j \sin(t\sigma y)) e^{-\frac{y^2}{2}} dy \\ &= \frac{2}{\sqrt{2\pi}} e^{jtm} \int_{-\infty}^{\infty} \cos(t\sigma y) e^{-\frac{y^2}{2}} dy \\ &= \frac{2}{\sqrt{2\pi}} e^{jtm} \frac{\sqrt{\pi} e^{-\frac{t^2\sigma^2}{2}}}{\sqrt{2}} \\ &= e^{\left(jtm - \frac{t^2\sigma^2}{2} \right)} \end{aligned}$$

Differentiation of this form will yield correctly the first 2 moments of the distribution.

Most important property of normal variables: any linear combination (weighted sum) of normal variables, whether independent or not, is another normal variable.

Note that for zero mean variables

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$
$$\phi(t) = e^{-\frac{\sigma^2 t^2}{2}}$$



Both are Gaussian forms.

The Normal Approximation to the Binomial Distribution

The binomial distribution deals with the outcomes of n independent trials of an experiment. Thus if n is large, we should expect the binomial distribution to be well approximated by the normal distribution. The approximation is given by the normal distribution having the same mean and variance. Thus

$$b(k, n, p) \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}$$

Relative error is of the order of $\frac{(k-np)^3}{(npq)^2}$

The relative fit is good near the mean if npq is large and degenerates in the tails where the probability itself is small.

The Normal Approximation to the Poisson Distribution

Also the Poisson distribution depends on the outcomes of independent events. If there are enough of them,

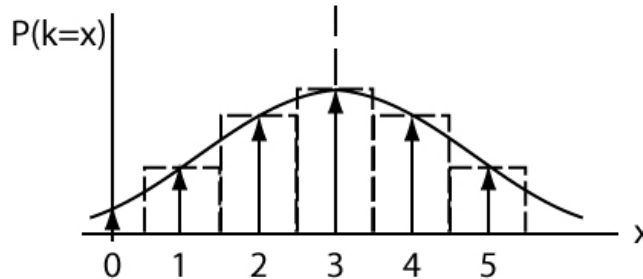
$$P(k, \mu) \approx \frac{1}{\sqrt{2\pi\mu}} e^{-\frac{(k-\mu)^2}{2\mu}}$$

Relative error is of the order of $\frac{(k-\mu)^3}{\mu^2}$

The relative fit is subject to the same behavior as the binomial approximation.

Interpretation of a continuous distribution approximating a discrete one:

The value of the normal density function at any k approximates the value of the discrete distribution for that value of k . Think of spreading the area of each impulse over a unit interval. Then the height of each rectangle is the probability that the corresponding value of k will be taken. The normal curve approximates this step-wise function.



Note that in summing the probabilities for values of k in some interval, the approximating normal curve should be integrated over that interval plus $\frac{1}{2}$ on each end to get all the probability associated with those values of k .

$$P(N_1 \leq X \leq N_2) = \sum_{k=N_1}^{N_2} P(k)$$

$$\sum_{k=N_1}^{N_2} P(k, \mu) \approx \int_{N_1 - \frac{1}{2}}^{N_2 + \frac{1}{2}} \frac{1}{\sqrt{2\pi\mu}} e^{-\frac{(x-\mu)^2}{2\mu}} dx$$

$$= \Phi\left(\frac{N_2 + \frac{1}{2} - \mu}{\sqrt{\mu}}\right) - \Phi\left(\frac{N_1 - \frac{1}{2} - \mu}{\sqrt{\mu}}\right)$$

Multidimensional Normal Distribution

Probability density function:

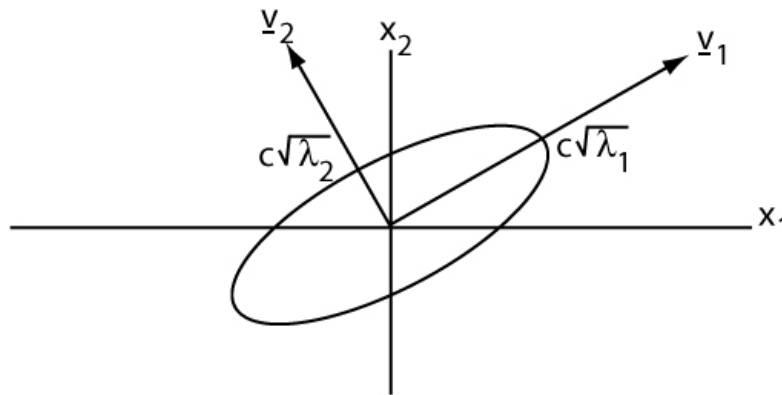
$$f(\underline{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{|M|}} \exp\left[-\frac{1}{2}(\underline{x} - \bar{X})^T M^{-1}(\underline{x} - \bar{X})\right]$$

Assuming zero mean, which is often the case:

$$f(\underline{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{|M|}} \exp\left[-\frac{1}{2} \underline{x}^T M^{-1} \underline{x}\right]$$

For zero mean variables: contours of constant probability density are given by:

$$\underline{x}^T M^{-1} \underline{x} = c^2$$



Not expressed in principal coordinates if the X_i are correlated.

Need to know the rudimentary properties of eigenvalues and eigenvectors.

M is symmetric and full rank.

$$M \underline{v}_i = \lambda_i \underline{v}_i$$

$$\underline{v}_i^T \underline{v}_j = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

This probability density function can be better visualized in terms of its principal coordinates. These coordinates are defined by the directions of the eigenvectors of the covariance matrix. The appropriate transformation is

$$\underline{y} = V \underline{x}$$

$$V = \begin{bmatrix} \leftarrow \underline{v}_1^T \rightarrow \\ \vdots \\ \leftarrow \underline{v}_n^T \rightarrow \end{bmatrix}$$

Thus y_i is the component of \underline{x} in the direction \underline{v}_i .

(In terms of the new variable \underline{y} , the contours of constant probability density are)

$$\begin{aligned}
 \underline{x}^T M^{-1} \underline{x} &= \underline{y}^T V^{-T} M^{-1} V^{-1} \underline{y} \\
 &= \underline{y}^T Y^{-1} \underline{y} \\
 Y^{-1} &= V^{-T} M^{-1} V^{-1} \\
 Y &= V M V^T \\
 &= V M \begin{bmatrix} \uparrow & \uparrow \\ \underline{v}_1 & \dots & \underline{v}_n \\ \downarrow & & \downarrow \end{bmatrix} \\
 &= \begin{bmatrix} \leftarrow \underline{v}_1^T \rightarrow \\ \dots \\ \leftarrow \underline{v}_n^T \rightarrow \end{bmatrix} \begin{bmatrix} \uparrow & \uparrow \\ \lambda_1 \underline{v}_1 & \dots & \lambda_n \underline{v}_n \\ \downarrow & & \downarrow \end{bmatrix} \\
 &= \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}
 \end{aligned}$$

Y is the covariance matrix for the random variable $\underline{Y} = V\underline{X}$, so the λ_i are the variance of the Y_i .

$$Y^{-1} = \begin{bmatrix} \frac{1}{\lambda_1} & & \\ & \ddots & \\ & & \frac{1}{\lambda_n} \end{bmatrix}$$

$$\underline{y}^T \begin{bmatrix} \frac{1}{\lambda_1} & & \\ & \ddots & \\ & & \frac{1}{\lambda_n} \end{bmatrix} \underline{y} = c^2$$

$$\underline{y}^T Y^{-1} \underline{y} = \frac{y_1^2}{\lambda_1} + \frac{y_2^2}{\lambda_2} + \dots + \frac{y_n^2}{\lambda_n} = c^2$$

These are the principal coordinate with intercepts at $y_i = \pm c\sqrt{\lambda_i}$ with $\sqrt{\lambda_i}$ the standard deviation of y_i .

Note that two random variables, each having a normal distribution singly, do not necessarily have a binormal joint distribution.

However, if the random variables are independent and normally distributed, their joint distribution is clearly a multidimensional normal distribution.

Two dimensional case- continued:

$$m_{ij} = \overline{X_i X_j}$$

$$f_2(x_1, x_2) = \frac{1}{2\pi\sqrt{m_{11}m_{22} - m_{12}^2}} \exp\left[-\frac{m_{22}x_1^2 - 2m_{12}x_1x_2 + m_{11}x_2^2}{2(m_{11}m_{22} - m_{12}^2)}\right]$$

$$m_{11} = \sigma_1^2$$

$$m_{22} = \sigma_2^2$$

$$m_{12} = \mu_{12}$$

$$\rho_{12} = \frac{\mu_{12}}{\sigma_x \sigma_y}$$

In terms of these symbols:

$$f_2(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}} \exp\left[-\frac{\left(\frac{x_1}{\sigma_1}\right)^2 - 2\rho_{12}\left(\frac{x_1}{\sigma_1}\right)\left(\frac{x_2}{\sigma_2}\right) + \left(\frac{x_2}{\sigma_2}\right)^2}{2(1-\rho_{12}^2)}\right]$$

Note that if a set of random variables having the multidimensional normal distribution is uncorrelated, they are independent. This is not true in general.