Probability & Statistics.

'C/C'

Class test - 9th April (9:30 - 11:10 am and 11:30 am - 1:10 pm)

20 marks

Syllabus - after midsem — till today.

Central limit theorem (CLT)

$X_1, X_2, \ldots$ a sequence of independent, identically distributed random variables, each with mean μ and variance σ^2.

Then

$$P\left\{ \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \le a \right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2}\, dx$$

as $n \to \infty$.

Prob The lifetime of a special type of battery is a random variable with mean 40 hrs and S.D. 20 hrs. A battery is used until it fails, at which point it is replaced by a new one. Assuming a stockpile of 25 such batteries the lifetimes of which are independent, approximate the probability that over 1100 hrs of use can be obtained.

Sol^n. X_i – The lifetime of the ith battery to be put in use, then we need to calculate

$$P\{X_1 + X_2 + \cdots + X_{25} > 1100\}.$$

$$= P\left\{\frac{X_1 + \cdots + X_{25} - 1000}{20\sqrt{25}} > \frac{1100 - 1000}{20\sqrt{25}}\right\}$$

$$\simeq P\{N(0,1) > 1\}$$

$$= 1 - \Phi(1) = 0.1587.$$

Proof of CLT.

Suppose that X_i have mean 0 and variance 1.

Then the MGF of $\dfrac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}$ is

$$\psi(t) = E\left[\exp\left\{t \frac{X_1 + \cdots + X_n}{\sqrt{n}}\right\}\right]$$

$$= E\left[e^{tX_1/\sqrt{n}} \cdot e^{tX_2/\sqrt{n}} \cdots e^{tX_n/\sqrt{n}}\right]$$

$$= \left(E\left[e^{tX/\sqrt{n}}\right]\right)^n$$

Now for large 'n'

$$e^{tX/\sqrt{n}} = 1 + \frac{tX}{\sqrt{n}} + \frac{t^2X^2}{2n} + O(\cdot)$$

$$\therefore E\left[e^{tX/\sqrt{n}}\right] = 1 + E\left[\frac{tX}{\sqrt{n}}\right] + E\left[\frac{t^2X^2}{2n}\right]$$

$$= 1 + t/\sqrt{n}\, E[X] + t^2/2n\, E[X^2]$$

$$= 1 + \frac{t^2}{2n}$$

Thus when n is large

$$E\left[\exp\left\{t\frac{X_1+X_2+\cdots+X_n}{\sqrt{n}}\right\}\right]$$

$$\simeq \left(1+\frac{t^2}{2n}\right)^n$$

If $n\to\infty$ the approximation can be shown to become exact so we have

$$\lim_{n\to\infty} E\left[\exp\left\{t\frac{X_1+\cdots+X_n}{\sqrt{n}}\right\}\right] = e^{t^2/2}$$

the MGF of $N(0,1)$

Thus the MGF of $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ converges to the MGF of $N(0,1)$.

Using this it can be proven that the distribution of the r.v. $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ converges to the standard Normal distribution Φ.

When X_i have mean μ and variance σ^2, the random variables $Y_i = \frac{X_i - \mu}{\sigma}$ have mean 0 and variance σ.

$$\frac{Y_1 + Y_2 + \cdots + Y_n}{\sqrt{n}} \to \Phi$$

Thus, the preceding shows that

$$P\left\{ \frac{X_1 - \mu + X_2 - \mu + \cdots + X_n - \mu}{\sigma\sqrt{n}} \le a \right\} \longrightarrow \Phi(a).$$

This proves the CLT.

Let us ensure that the MGF of $N(0,1)$ is $e^{t/2}$!!

Let us calculate MGF of $N(\mu, \sigma^2) = X$.

$$\psi(t) = E[e^{tX}]$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[\frac{-[x^2 - 2(\mu+\sigma^2 t)x + \mu^2]}{2\sigma^2}\right] dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{2\mu\sigma^2 t + \sigma^4 t^2}{2\sigma^2}\right) \int_{-\infty}^{\infty} \exp\left[\frac{-[x-(\mu+\sigma^2 t)]^2}{2\sigma^2}\right] dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right) \int_{-\infty}^{\infty} e^{-y^2} (\sigma\sqrt{2})\, dy$$

1 //

(H.W)

$$y = \frac{x - (\mu + \sigma^2 t)}{\sigma\sqrt{2}}$$

$$\Rightarrow dy = dx/(\sigma\sqrt{2})$$

$$= \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$$

Putting $\mu = 0, \sigma^2 = 1$,

we obtain the MGF of $N(0,1)$ as $e^{t/2}$.

Bivariate normal distribution.

Q. Do one-dimensional normal distribution and the one-dimensional CLT allow for a generalization to dimension two or higher?

A. Yes.

A random vector $\begin{bmatrix} X \\ Y \end{bmatrix}$ is said to have a '<u>standard bivariate normal distribution</u>' <u>with parameter</u> ρ if it has a joint probability density of the form.

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2}(x^2-2\rho xy+y^2)/(1-\rho^2)}$$

$$-\infty < x, y < \infty$$

and $-1 < \rho < 1$.

We will show that ρ is the ~~Cov~~ correlation coeff. of X & Y.

First find out the marginal distribution ~~of~~ of X & Y.

$$f(x,y) = \underline{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}} \cdot \underline{\frac{1}{\sqrt{1-\rho^2}\sqrt{2\pi}} e^{-\frac{1}{2}(y-\rho x)^2/(1-\rho^2)}}$$

For a fixed x,

$$g(y) = \frac{1}{\sqrt{1-\rho^2}\sqrt{2\pi}} e^{-\frac{1}{2}(y-\rho x)^2/(1-\rho^2)}$$

is an $N(\rho x, 1-\rho^2)$ density.

This implies $\int_{-\infty}^{\infty} g(y)\,dy = 1$

So

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad -\infty < x < \infty$$

The marginal density $f_X(x)$ of X is the standard normal density.

Also, due to symmetry, the marginal density $f_Y(y)$ of Y is the standard normal ~~normal~~ density.

Now we prove that $\rho = \text{Corr}(X,Y)$.

Since $\text{Var}(X) = \text{Var}(Y) = 1$

it is enough to show that

$$\text{Cov}(X,Y) = \rho.$$

Note that $E[X] = 0 = E[Y]$

and $\text{Cov}(X,Y) = E[XY]$

$$\therefore \text{Cov}(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy \, f(x,y)\,dx\,dy$$

$$= \int_{x=-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \, dx \int_{y=-\infty}^{\infty} y \frac{1}{\tau\sqrt{2\pi}} e^{-\frac{1}{2}(y-\rho x)^2/\tau^2} \, dy$$

where $\tau^2 = 1-\rho^2$

$$= \int_{-\infty}^{\infty} \rho x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \, dx$$

$\qquad\qquad || \quad E[Y]$ where $Y \sim N(\rho x, \tau^2)$

$$= \rho E[X^2] = \rho$$

$X \sim N(0,1)$

General version.

A random vector (X, Y) is said to be bivariate normally distributed with parameters $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ if the standardized random vector

$$\left(\frac{X - \mu_1}{\sigma_1}, \frac{Y - \mu_2}{\sigma_2} \right)$$

has the standard bivariate normal distribution with parameter ρ.

The joint density $f(x,y)$ of the random variables X & Y is given by

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right) - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right] / (1-\rho^2)}$$

Conditional distribution

$$f_{X_2|X_1}(x_2|x_1) \text{ and }$$

$$f_{X_1|X_2}(x_1|x_2)$$

$$f_{X_1} \sim N(\mu_1, \sigma_1)$$

$$f_{X_2|X_1}(x_2|x_1) = \frac{f(x_1, x_2)}{f_{X_1}(x_1)}$$

this should be sigma2/sigma1 ✓

$$= \frac{1}{\sigma_2\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2\sigma_2^2(1-\rho^2)}\left\{x_2 - \left[\mu_2 + \rho\sigma_1\sigma_2(x_1-\mu_1)\right]\right\}^2\right]$$

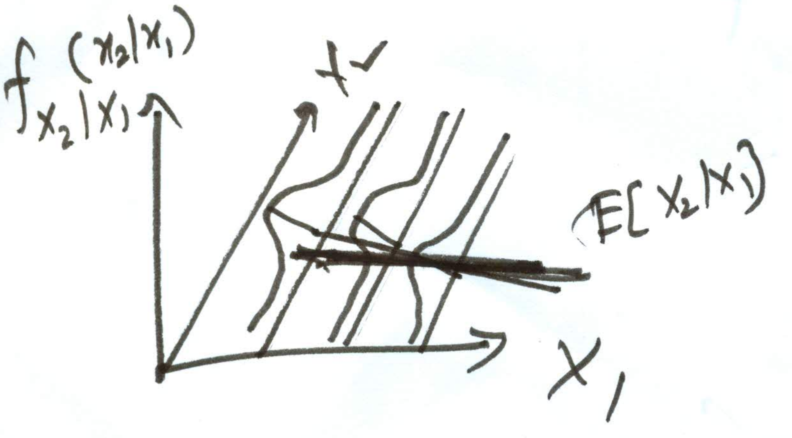
$$-\infty < x_2 < \infty$$

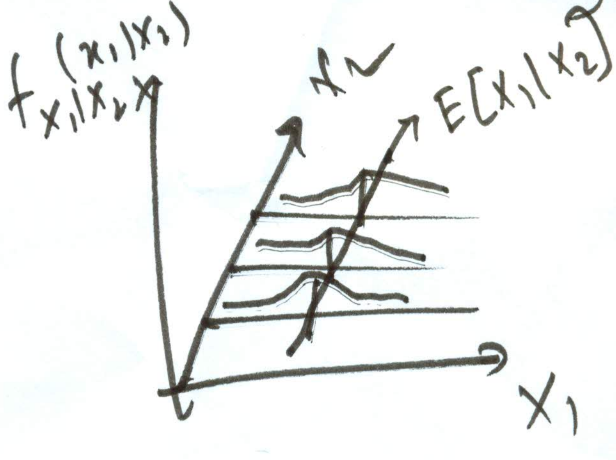
and

$$f_{X_1|X_2}(x_1|x_2) = \frac{f(x_1, x_2)}{f_{X_2}(x_2)}$$

this should be sigma1/sigma2 ✓

$$= \frac{1}{\sigma_1\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2\sigma_1^2(1-\rho^2)}\left\{x_1 - \left[\mu_1 + \rho\sigma_2\sigma_1(x_2-\mu_2)\right]\right\}^2\right]$$





Observe that

$$X_2 \mid X_1 \sim N\left[\underline{\mu_2 + \rho\left(\sigma_2/\sigma_1\right)(x_1 - \mu_1)}, \underline{\sigma_2^2(1-\rho^2)}\right]$$

and

$$X_1 \mid X_2 \sim N\left[\mu_1 + \rho\left(\sigma_1/\sigma_2\right)(x_2 - \mu_2), \underline{\sigma_1^2(1-\rho^2)}\right]$$

$$E[X_2 \mid X_1] = \cancel{\mu_2}$$

$$E(X_1 \mid X_2) =$$

Log-normal distribution

(H.W)

part of the syll.