Probability & Statistics.

L-17

Central limit theorem (CLT)

$\underbrace{X_1, X_2, \ldots}_{\text{iid}}$ RVs with mean μ, variance σ^2

$E(X^2) < \infty$

$$S_n = X_1 + X_2 + \cdots + X_n$$

$$P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right) \approx \Phi(x), \quad -\infty < x < \infty$$

$$\Rightarrow \quad P(S_n \le x) \approx \Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right)$$

$$= \Phi\left(\frac{x - E(S_n)}{\sqrt{Var(X_n)}}\right)$$

very very useful.

normal approximation formula

Bivariate Normal Distribution.

Q. Do ~~not~~ one-dimensional normal distribution and the one-dimensional CLT allow for a generalization to dimension two or higher?

Ans Yes.
As the 1-d normal density is completely determined by its expected value and variance, the bivariate normal density is completely specified by the expected values and the variances of its marginal densities and correlation coefficient.

Bivariate Standard normal.

A random vector (X, Y) is said to have a standard bivariate normal distribution with parameter ρ if it has a joint probability density fn. of the form

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

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

$$\rho = \rho(X, Y)$$

Note that $\qquad = \text{the correlation coeff.}$

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

Now for any fixed x,

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

$$\approx N(\rho x, 1-\rho^2)$$

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

[...]erefore,

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad -\infty < x < \infty$$

$$\approx N(0,1).$$

[...]ng the symmetry in $f(x,y)$, $f_Y(y) \approx N(0,1)$.

proof $\rho(X,Y) = \rho$.

Note that $V(X) = 1 = V(Y)$

$E(X) = 0 = E(Y)$.

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

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)\,dx\,dy.$$

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

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

$$= \left(\int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \right) \cdot \rho x$$

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

$$= \rho\, E(X^2)$$

$$= 1\rho.$$

General form of Bivariate Normal Distribution.

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 bivariate standard normal distribution.

$\therefore$ the joint density $f(x,y)$ of X, Y is given by

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

Observation:

1. The marginal densities $f_X(x)$ & $f_Y(y)$ of X & Y are $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ respectively.
2. $\rho = \rho(X, Y)$.

Note: In general, uncorrelated is a necessary but not sufficient cond^n for independence of two random variables.

However, for a bivariate normal distribution, uncorrelated is a necessary & suff. cond^n for independence.

— just set $\rho = 0$ in $f(x, y)$.

The conditional distribution $f_{Y|X}(y|x)$ and $f_{X|Y}(x|y)$

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}$$

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

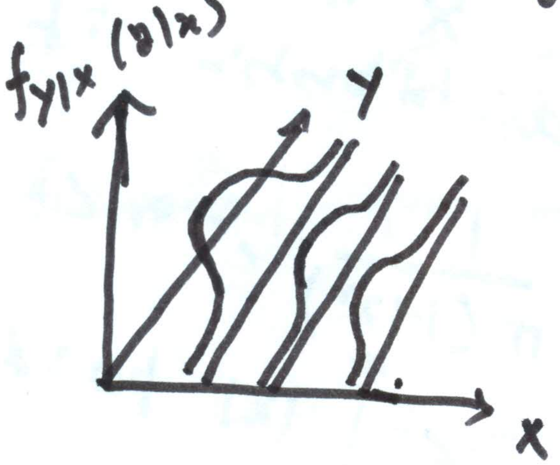
$$\sim N\left[\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1),\ \sigma_2^2(1-\rho^2)\right] -$$

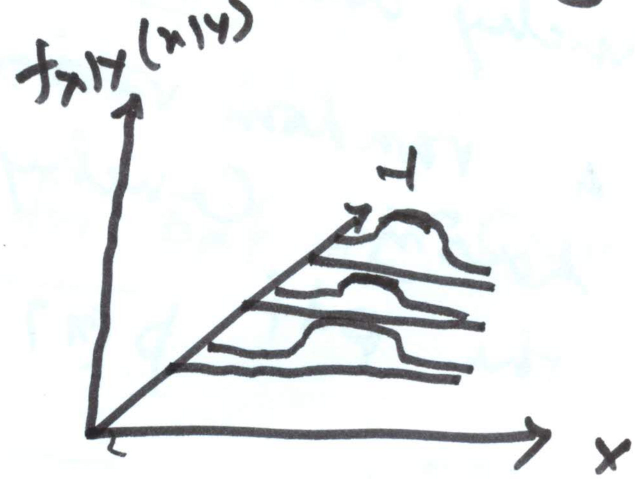
Check!!

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$$

$$= \frac{1}{\sigma_1\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left[\ \cdot \quad \cdot \quad \cdot\ \right]$$

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





Log normal distribution.

$X(\omega) > 0$

Consider a random variable X with the range $\{x \mid 0 < x < \infty\}$ such that $Y = \ln X$ is normally distributed with mean μ_Y & variance σ_Y^2.

Then the density of X is

$$f(x) = \begin{cases} \frac{1}{x\sigma_Y\sqrt{2\pi}} e^{-\frac{1}{2}\left[(\ln x - \mu_Y)/\sigma_Y\right]^2}, & x > 0 \\ 0, & \text{otherwise.} \end{cases}$$

$$E(X) = e^{\mu_y + \frac{1}{2}\sigma_y^2}$$

$$V(X) = \mu_x^2 \left(e^{\sigma_y^2} - 1 \right).$$

Cauchy distribution.

A random variable X is said to having Cauchy distribution if the pdf

$$p(x) = \frac{1}{\pi(1+x^2)}, \quad -\alpha < x < \alpha$$

$$E(X) = \int_{-\alpha}^{\infty} x \, p(x) \, dx \leq \int_{-\infty}^{\infty} |x| \, p(x) \, dx.$$

Now $\int_{-\infty}^{\infty} |x| \, p(x) \, dx = \int_{-\infty}^{\infty} \frac{1}{\pi} \frac{|x|}{1+x^2} \, dx$

$$= \int_{0}^{\infty} \frac{2}{\pi} \frac{x}{1+x^2} \, dx.$$

$$\frac{x}{1+x^2} > \frac{1}{2x} \quad \forall \, x > 1$$

$$\int_{-\alpha}^{\infty} \frac{1}{\pi} \frac{|x|}{1+x^2} \, dx > 2 \int_{1}^{\infty} \frac{1}{\pi} \frac{1}{x} \, dx$$

$$= \frac{2}{\pi} \log(\infty) = \alpha.$$

$$E(x) = \int_{-\infty}^{0} x\, p(x)\, dx + \int_{0}^{\infty} x\, p(x)\, dx$$

$$= -\infty + \infty \rightarrow \underline{\text{undefined}}.$$

$E(x)$, $V(x)$, and $E(x^n)$ $\forall n$ do not exist.

General <u>formula for</u> pdf of Cauchy.

$$p(x) = \frac{1}{\alpha\pi\left[1 + \frac{(x-a)^2}{\alpha^2}\right]}, \quad -\infty < x < \infty$$

look at the webpage (BW) for justification.

a — scale parameter

α — location parameter

Weibull distribution.

The pdf of Weibull distribution:

$$f(x;\lambda,k)=\begin{cases}\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1}e^{-\left(\frac{x}{\lambda}\right)^{k}}, & x\geq 0\\ 0, & x<0\end{cases}$$

where $k>0$ is the shape parameter
$\lambda>0$ is the scale parameter.

If the quantity X is 'time to failure' the Weibull dist. gives a distribution for which failure rate is proportional to the power of time.

The shape parameter k, is that power $+1$.

∴ (1) $k<1$ indicates the failure rate decreases over time.

(2) $k=1$ indicates the failure rate is constant over time

(3) $k<1$, the failure rate increases over time.

$$E(X)=\lambda^{n}\ \Gamma\left(1+\frac{n}{k}\right),\ V(X)=\lambda^{2}\left[\Gamma\left(1+\frac{2}{k}\right)-\left(\Gamma\left(1+\frac{1}{k}\right)\right)^{2}\right]$$