Cheat Sheet on Beta & Gamma function

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$$

Recurrence formula

$$\boxed{\Gamma(z+1) = z\Gamma(z)}$$

$$\Gamma(1) = 1 \qquad \Gamma(n+1) = n!$$

for n positive integer

Reflection formula

$$\boxed{\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin\pi z}} \Rightarrow \Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi}$$

Stirling $\quad \Gamma(z+1) \sim \sqrt{2\pi z}\, e^{-z} z^{z}$

for large z

$$B(a,b) = \int_0^1 dx\, x^{a-1}(1-x)^{b-1} dx$$

$x = \frac{t}{1+t}$

$$= \int_0^\infty \frac{t^{a-1}}{(1+t)^{a+b}} dt$$

$x = \sin^2\theta$

$$= 2\int_0^{\pi/2} \sin^{2a-1}\theta \cos^{2b-1}\theta\, d\theta$$

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

That is the probability density function of the standard normal variate

In general

Normal distribution:

Pdf

$$\mathcal{N}(x|\mu,\sigma^2) \triangleq \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$E[X] = \mu, \quad \text{var}[X] = \sigma^2$$

We will often use $X \sim \mathcal{N}(\mu,\sigma^2)$

Cumulative distribution:

$$\Phi(x;\mu,\sigma^2) \triangleq \int_{-\infty}^{x} \mathcal{N}(z|\mu,\sigma^2)\,dz$$

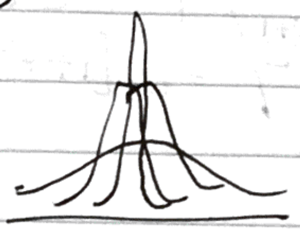
$$= \frac{1}{2}\; 1+\text{erf}(z/\sqrt{2})$$

where $\text{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt$.

Few other continuous distributions

Degenerate pdf / delta function

$$\lim_{\sigma \to 0} \mathcal{N}(x|\mu, \sigma^2) = \delta(x-\mu)$$



$$\int dx\, f(x)\, \delta(x-\mu) = f(\mu)$$

The Student's t distribution

$$\mathcal{T}(x|\underset{\text{location}}{\mu}, \underset{\text{scale/dispersion}}{\sigma^2}, \nu) \propto \left[1 + \frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right]^{-\frac{\nu+1}{2}}$$

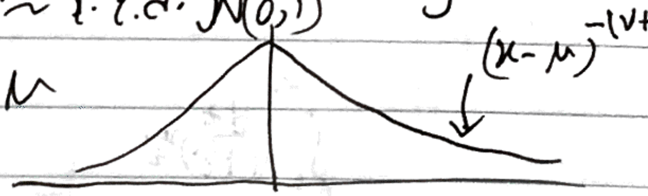
ν is called the degree of freedom.

The distribution with $\mu = 0$, $\sigma = 1$ plays an important role in hypothesis testing.

$X \sim z / \sqrt{\left[\sum_{i=1}^{\nu} z_i^2 / \nu\right]}$ where $z, z_i \sim$ i.i.d. $\mathcal{N}(0,1)$

Symmetric around μ

$E[X] = \mu$



Power law tail. More tolerant to outliers

σ^2 not variance. For $\nu \leq 2$ $Var[X] = \infty$

For $\nu > 2$

$$\mathrm{Var}[X] = \int_{-\infty}^{\infty} \frac{(x-\mu)^2\, dx}{\left[1+\frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right]^{\frac{\nu+1}{2}}} \Big/ \int_{-\infty}^{\infty} \frac{dx}{\left[1+\frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right]^{\frac{\nu+1}{2}}}$$

$$= \nu\sigma^2 \int_0^\infty \frac{z^2\, dz}{[1+z^2]^{\frac{\nu+1}{2}}} \Big/ \int_0^\infty \frac{dz}{[1+z^2]^{\frac{\nu+1}{2}}}$$

$$z = \frac{1}{\sqrt{\nu}} \frac{x-\mu}{\sigma}$$

$$= \nu\sigma^2 \frac{\int_0^\infty \frac{1+z^2}{(1+z^2)^{\frac{\nu+1}{2}}}\, dz - \int_0^\infty \frac{dz}{(1+z^2)^{\frac{\nu+1}{2}}}}{\int_0^\infty \frac{dz}{(1+z^2)^{\frac{\nu+1}{2}}}} = \nu\sigma^2 \left[\frac{\int_0^\infty \frac{dz}{(1+z^2)^{\frac{\nu-1}{2}}}}{\int_0^\infty \frac{dz}{(1+z^2)^{\frac{\nu+1}{2}}}} - 1 \right]$$

Use $z = \sqrt{t}$

$$\int_0^\infty \frac{dz}{(1+z^2)^{\frac{\rho+1}{2}}} = \frac{1}{2} \int_0^\infty \frac{t^{-1/2}\, dt}{(1+t)^{\frac{\rho+1}{2}}} = \frac{1}{2} B\left(\frac{1}{2}, \frac{\rho}{2}\right)$$

$$= \frac{1}{2} \frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\rho}{2}\right)}{\Gamma\left(\frac{\rho+1}{2}\right)}$$

Since $B(x,y) = \int_0^\infty \frac{t^{x-1}}{(1+t)^{x+y}}\, dt$

$$\frac{\int_0^\infty \frac{dz}{(1+z^2)^{\frac{\nu-1}{2}}}}{\int_0^\infty \frac{dz}{(1+z^2)^{\frac{\nu+1}{2}}}} = \frac{\left(\frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\nu-2}{2}\right)}{\Gamma\left(\frac{\nu-1}{2}\right)}\right)}{\left(\frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)}\right)} = \frac{\Gamma\left(\frac{\nu-2}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu-1}{2}\right)}$$

$$= \frac{\Gamma\left(\frac{\nu}{2}-1\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu-1}{2}\right)}$$

$x\Gamma(x) = \Gamma(x+1)$ So $\Gamma\left(\frac{\nu}{2}\right) = \left(\frac{\nu}{2}-1\right)\Gamma\left(\frac{\nu}{2}-1\right)$

$$= \frac{\nu-2}{2}\Gamma\left(\frac{\nu}{2}-1\right)$$

and $\Gamma\left(\frac{\nu+1}{2}\right) = \frac{\nu-1}{2}\Gamma\left(\frac{\nu-1}{2}\right)$

$$\operatorname{Var} X = \nu\sigma^2\left[\frac{\cancel{2}}{\nu-2}\,\frac{\nu-1}{\cancel{2}} - 1\right]$$

$$= \frac{\nu\sigma^2}{\nu-2}$$

$\nu = 1$ $\quad T(x|\mu,\sigma^2,\nu) \propto \dfrac{1}{1+\left(\frac{x-\mu}{\sigma}\right)^2}$

Cauchy distribution

Lorenz "

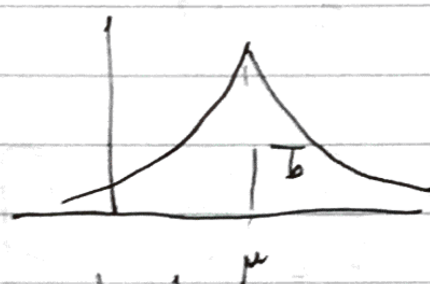
Note that $M_X(t) = E\left[e^{tX}\right]$ is ill defined for $t \in \mathbb{R}$.

~~Blowup~~ Higher moments diverge!

$\varphi_X(t) = E\left[e^{itX}\right]$ is defined.

Laplace distribution

$$\mathrm{Lap}(x|\mu, b) \triangleq \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right)$$



$E[x] = \mu$

$\mathrm{var}[X] = 2b^2$

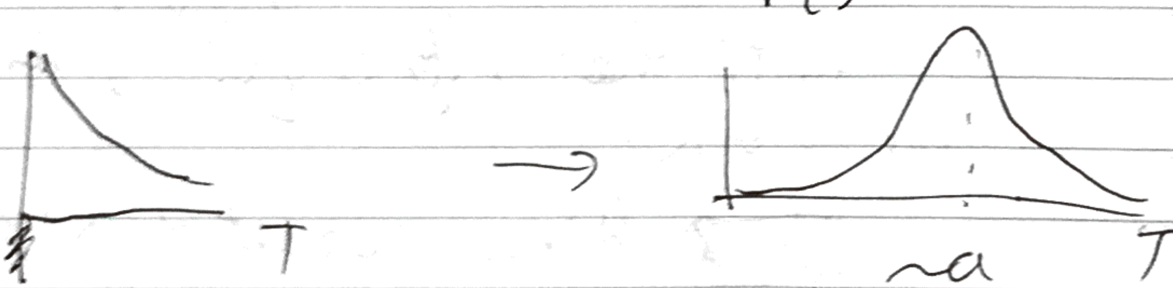
$$\int_{-\infty}^{\infty} (x-\mu)^2 \frac{e^{-\frac{|x-\mu|}{b}}}{2b} dx = \frac{b^2}{2}\int_{-\infty}^{\infty} z^2 e^{-|z|} dz = b^2 \int_0^{\infty} z^2 e^{-z} dz = 2b^2$$

$\Gamma(3)$

Often used as priors in sparse ~~reg~~ estimation problem.

The gamma distribution

$$\mathrm{Ga}(T|\text{shape}=a, \text{rate}=b) \triangleq \frac{b^a}{\Gamma(a)} T^{a-1} e^{-Tb}$$



Small a

large a

Special cases

$\text{Expon}(x|\lambda) = Ga(x|a=1, b=\lambda) = \lambda e^{-\lambda x}$

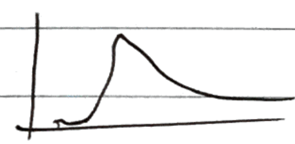
Waiting time in poisson process

Chi-squared $\quad \chi^2(x|\nu) = Ga(x|\frac{\nu}{2}, \frac{1}{2})$

$\propto x^{\frac{\nu}{2}-1} e^{-x/2}$

If $z_i \sim N(0,1)$ for $i=1,\dots,\nu$
and are indept.

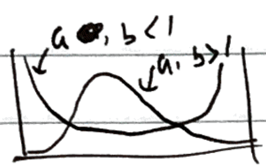
$S = \sum_{i=1} z_i^2 \qquad S \sim \chi_\nu$



Goodness of fit tests

Beta distribution

$\text{Beta}(x|a,b) = \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1}$



$E(X) = \frac{B(a+1,b)}{B(a,b)} = \frac{\Gamma(a+1)}{\Gamma(a)} \frac{\Gamma(a+b)}{\Gamma(a+b+1)}$

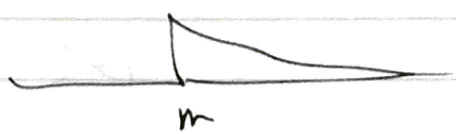
$= \frac{a}{a+b}$

etc.

Useful for ~~powers~~ priors in binomial / multinomial distribution

Pareto

$$\text{Pareto}(x|k,m) = k m^k x^{-(k+1)} \mathbb{I}(x \geq m)$$

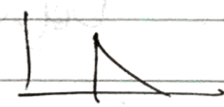


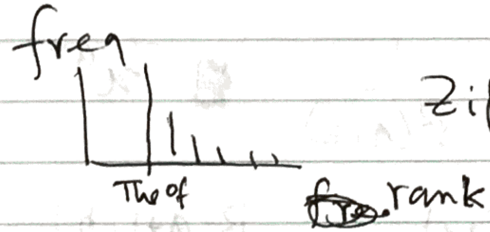
Long tailed distribution.

~~[cheat sheet]~~

$$E[X] = \frac{km}{k-1}$$

For large k, $E[x]$ is close to m





Zipf's law

Cheat Sheet on Beta & Gamma function

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, dt$$

Recurrence formula

$$\boxed{\Gamma(z+1) = z\,\Gamma(z)}$$

$\Gamma(1) = 1$ $\qquad\qquad$ $\Gamma(n+1) = n!$

for n positive integer

Reflection formula

$$\boxed{\Gamma(z)\,\Gamma(1-z) = \frac{\pi}{\sin \pi z}} \Rightarrow \Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi}$$

Stirling $\qquad \Gamma(z+1) \sim \sqrt{2\pi z}\; e^{-z} z^{z}$

for large z

$$B(a,b) = \int_0^1 \cancel{dx}\; x^{a-1}(1-x)^{b-1}\, dx$$

$x = \frac{t}{1+t}$

$$= \int_0^\infty \frac{t^{a-1}}{(1+t)^{a+b}}\, dt$$

$x = \sin^2\theta$

$$= 2\int_0^{\pi/2} \sin^{2a-1}\theta \,\cos^{2b-1}\theta \, d\theta$$

$$B(a,b) = \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}$$

Joint ~~Probability~~ prob. distr^n

$p(x_1, \dots, x_D)$

Covariance & Correlation

$$Cov[X,Y] = E[(X - E[X])(Y - E[Y])]$$

If $\mathbf{x} \in \mathbb{R}^d$ covariance ~~is~~ a matrix.

$$Cov[\mathbf{x}] = \begin{pmatrix} Var[X_1] & Cov[X_1, X_2] & \cdots \\ Cov[X_1, X_2] & Var[X_2] & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

$$Corr[X,Y] = \frac{Cov[X,Y]}{Var[X]\,Var[Y]}$$

$$-1 \le corr[X,Y] \le 1$$

Essentially a ~~result~~ consequence of $\left|\sum_i x_i y_i\right| \le \left(\sum_i x_i^2\right)^{1/2} \left(\sum_i y_i^2\right)^{1/2}$

$Cov[X,Y] = 1$ iff $Y = aX + b$

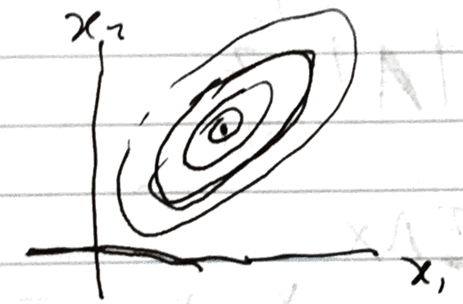
If X, Y are indept $Cov[X,Y] = 0$
but not vice versa!

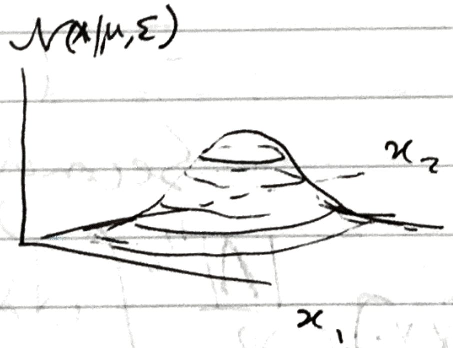
The multivariate Gaussian

$$\mathcal{N}(x|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{D/2}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$$

μ mean vector.
Σ covariance matrix

$\Lambda = \Sigma^{-1}$ called precision matrix
or concentration matrix.





Few things:

$$\int d^Dx \, e^{-\frac{1}{2}x^TAx} = (2\pi)^{D/2}|A|^{-1/2}$$

determinant or A

Easy way to ~~to~~ see: diagonalize $A = O^T A_D O$

$$\Lambda_O = \begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_D\end{pmatrix}$$

Let $Ox = z$. $\quad d^Dx \to d^Dz$ since $|O| = 1$

$$\int d^Dx \, e^{-\frac{1}{2}x^T\Lambda x}$$

$$= \int d^Dz \, e^{-\frac{1}{2}z^T\Lambda_O z} = \prod_{i=1}^{D}\left(\int dz_i \, e^{-\frac{1}{2}z_i^2}\right)$$

$$= \prod_{i=1}^{D}\sqrt{2\pi/\lambda_i} = (2\pi)^{D/2}\left(\prod_i \lambda_i\right)^{-1/2}$$

$$= (2\pi)^{D/2}|A|^{-1/2}$$

Second

$$E[x_i x_j] = \frac{|\Lambda|^{1/2}}{(2\pi)^{D/2}}\int d^Dx \, e^{-\frac{1}{2}x^T\Lambda x} x_i x_j = (\Lambda^{-1})_{ij}$$

$$E\left[e^{t^T x}\right] = C\int d^Dx \, e^{-\frac{1}{2}x^T\Lambda x + t\cdot x}$$

$$M_X[t] = E\left[e^{\sum_i t_i x_i}\right]$$ could be used for moment calculation

$$\frac{1}{2}x^T \Lambda x - t \cdot x$$

$$= \frac{1}{2}(x - \Lambda^{-1} t)^T \Lambda (x - \Lambda^{-1} t)$$

$$- \frac{1}{2} t^T \Lambda^{-1} t$$

$$E\left[e^{t^T x}\right] = c \int d^D x \, e^{-\frac{1}{2}(x - \Lambda^{-1} t)^T \Lambda (x - \Lambda^{-1} t)}$$

$$\times e^{\frac{1}{2} t^T \Lambda^{-1} t}$$

Changing varible to $y = x - \Lambda^{-1} t$

$$\underbrace{\frac{|\Lambda|^{1/2}}{(2\pi)^{D/2}} \int d^D y \, e^{-\frac{1}{2} y^T \Lambda y}}_{1} \times e^{\frac{1}{2} t^T \Lambda^{-1} t}$$

$$\frac{\partial}{\partial t_i} \frac{\partial}{\partial t_j} M_x[t] \Big|_{t=0} = c \int d^D x \; x_i x_j e^{-\frac{1}{2} x^T \Lambda x + t^T x} \Big|_{t=0}$$

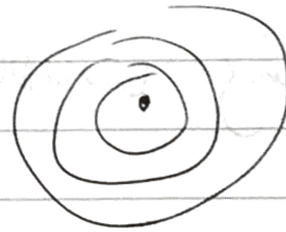
$$= E[X_i X_j]$$

$$\partial_{t_i}\partial_{t_j}\, e^{\frac{1}{2}\sum_{ij} t_i \Lambda^{-1}_{ij} t_j}\Big|_{t=0}$$

$$= \Lambda^{-1}_{ij}$$

Isotropic covariance $\Sigma = \sigma^2 I_D$

$$\frac{1}{(\sqrt{2\pi\sigma^2})^D} e^{-\frac{(x_i-\mu_i)^2}{2\sigma_{\Sigma}^2}}$$



Multivariate t distr^n.

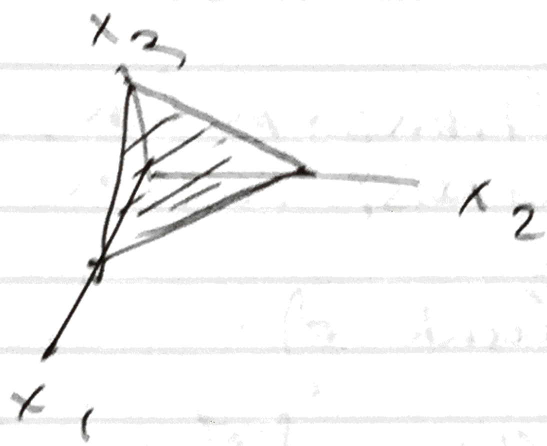
$$\tau(x|\mu,\Sigma,\nu) \propto \left[1+\frac{1}{\nu}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]^{-\frac{(\nu+D)}{2}}$$

Σ is not the covariance matrix

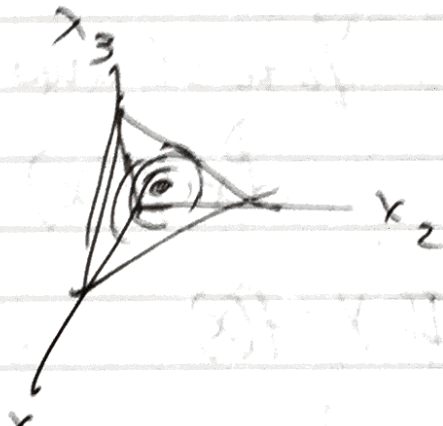
(Multivariate) Dirichlet distr^n.

$$Dir(x|\alpha) = \prod_k x_k^{\alpha_k - 1} \quad , 0 \leq x_k \leq 1$$
$$\sum x_k = 1$$

$\alpha_k = 1 \Rightarrow$ Uniform



α's high



Often used in Bayesian applic^ns

$x_i \to \theta_i$. Prior for multinomial distr^n parameters.

$$E[X_k] = \frac{\alpha_k}{\alpha_0} \qquad Var[X_k] = \frac{\alpha_k(\alpha_0 - \alpha_k)}{\alpha_0^2(\alpha_0+1)}$$

$$\alpha_0 = \sum \alpha_k$$

Coordinate transformation
an probability distribution

You should think of

$d^D x \, p(x)$ as a volume form

If $y = f(x)$ and $p_x(x) = p_y(f(x)) = p_y(y)$

$$d^D x \, p_x(x) = dx_1 \wedge dx_2 \wedge \cdots \wedge dx_D \, p_x(x)$$

$$\bigwedge^D \left(\sum_j \frac{\partial x_i}{\partial y_j} dy_j \right) = dy_1 \wedge \cdots \wedge dy_D \det\left(\frac{\partial x}{\partial y}\right)$$

$$d^D x \, p_x(x) = d^D y \det\left(\frac{\partial x}{\partial y}\right) p_y(y)$$

Special cases:

<u>1D linear transformation</u>

$y = ax + b$

$$dx\, p_x(x) = \frac{dy}{a}\, p_y(y) = \frac{1}{a}\, dy\, p_y(y)$$

<u>1D non-linear differentiable transform</u>

$$dx\, p_x(x) = \frac{dx}{dy}\, dy\, p_y(y)$$

<u>D dim linear</u>

$$y = A x + b$$

A: matrix, b: vector

A invertible.

One could use $A = O_y^T \Lambda O_x$

[Singular value decomposition]

Λ = diagon, $O_x, O_y \in O(D)$

So that $\quad v = \Lambda u + O_y^T b$

$\Lambda = (\lambda_1\ \lambda_2 \ldots)$

$$dv_i = \lambda_i du_i$$

$$d^D x = d^D u = \frac{1}{\lambda_1 \cdots \lambda_D}\, d^D v = \frac{1}{|A|}\, d^D v$$

2D non-linear ! $dx dy\, P_c(x,y) \rightarrow r dr d\theta\, P_p(r,\theta$

Expectation and ~~vo~~ Covariance under linear transformat

$$Y = AX + b$$

$$E[Y] = A\,E[X] + b$$

$$[\mathrm{Cov}(Y)]_{ij} = E\big[(Y_i - E[Y_i])(Y_j - E[Y_j])\big]$$

$$= \sum_{kl} A_{ik} A_{jl} E\big[(X_k - E[X_k])(X_l - E[X_l])\big]$$

$$= \left[A\,\mathrm{Cov}[X]\,A^T\right]_{ij}$$

$$\Sigma_Y = A \Sigma_X A^T$$

Before we go on to central limit theorem, note that for two independent variable X, Y, the sum $Z = X+Y$ has a characteristic function that is simply related to those of X and of Y

$$\varphi_Z(t) = E[e^{itZ}] = E[e^{itX}e^{itY}]$$
$$= E[e^{itX}]E[e^{itY}] = \varphi_X(t)\varphi_Y(t)$$

Because,

$$E[e^{itX}e^{itY}] = \sum_{x,y} p(x,\cdot)\,p(\cdot,y)\,e^{itx}e^{ity}$$

$$= \left(\sum_x p(x,\cdot)\,e^{itx}\right)\left(\sum_y p(\cdot,y)\,e^{ity}\right)$$

$$= E[e^{itX}]\;E[e^{itY}]$$

Same way $M_Z(t) = M_X(t)M_Y(t)$

For N independently and identically distributed variables X_i, if $S_N = \sum_{i=1}^{N} X_i$

$$\varphi_{S_N}(t) = \left[\varphi_{X_1}(t)\right]^N$$

One other thing: Cumulants

$$\varphi_X(t) = E\left[e^{itX}\right]$$
$$= \sum_{n=0}^{\infty} \frac{(it)^n}{n!} E[X^n] = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \mu_n$$

With moment generating function

$$M_X(t) = E\left[e^{tX}\right] = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mu_n$$

Now consider $K_X(t) = \ln M_X(t)$

In other words $M_X(t) = e^{K_X(t)}$

What is $K_X(t)$ like?

$$K_X(t) = \ln\left(1 + t\mu_1 + \frac{t^2}{2}\mu_2 + \frac{t^3}{3!}\mu_3 + \frac{t^4}{4!}\mu_4 \dots\right)$$
$$= \left(t\mu_1 + \frac{t^2}{2}\mu_2 + \frac{t^3}{3!}\mu_3 + \frac{t^4}{4!}\mu_4 \dots\right) - \frac{1}{2}\left(t\mu_1 + \frac{t^2}{2}\mu_2 + \frac{t^3}{3!} \dots\right)^2$$

$$+\frac{1}{3}\left(t\mu_1+\frac{t^2}{2}\mu_2\right)^3-\frac{1}{4}\left(t\mu_1+\cdots\right)^4+\cdots$$

$$=t\mu_1+\frac{t^2}{2}\left(\mu_2-\mu_1^2\right)+\frac{t^3}{3!}\left(\mu_3-3\mu_2\mu_1+2\mu_1^3\right)$$

$$+\frac{t^4}{4!}\left(\mu_4-4\mu_3\mu_1-3\mu_2^2+12\mu_2\mu_1^2-6\mu_1^4\right)$$

$$+\cdots$$

In general $K_x(t)=\sum_{n=1}^{\infty}\frac{t^n}{n!}\kappa_n$

If we define expectation of $(X-\mu)^n$ to be μ_n'

$\kappa_1=\mu_1=\mu \rightarrow$ mean

$\kappa_2=\mu_2-\mu_1^2=\sigma^2=\mu_2' \rightarrow$ variance

$\kappa_3=\mu_3-3\mu_2\mu_1+2\mu_1^3=\mu_3' \rightarrow$ Skewness

$\kappa_4=\mu_4-\cdots \quad =\mu_4'-3\mu_2'^2$

Kurtosis

Note that for normal distribution $M_x(t)=e^{t\mu+\frac{t\sigma^2}{2}} \Rightarrow \kappa_n=0$ for $n>2$

Note that if $Z = X+Y$, and X and Y are independent

$$M_Z(t) = M_X(t)\, M_Y(t)$$

Taking log

$$K_Z(t) = K_X(t) + K_Y(t)$$

So $$\kappa_n^{(z)} = \kappa_n^{(x)} + \kappa_n^{(y)}$$

For independent variables, cumulants add. We know this to be true of mean and variance, but it is true of higher cumulants as well.

Now we develop some intuition first. Let X_i be i.i.d variables with mean μ and variance σ^2, and $S_N = \sum_{i=1}^{N} X_i$ is their sum, then for this 'standardized'

Variable $Z_N = \frac{S_N - N\mu}{\sqrt{N}\,\sigma}$

$E[Z_N] = 0 \qquad \mathrm{var}[Z] = 1$

If X_i had higher cumulants κ_n for $n > 2$

$$\kappa_n^{(S_N)} = N\kappa_n^{(X_1)}$$

$$\kappa_n^{(S_N - N\mu)} = N\kappa_n^{(X_1)}$$

$$\kappa_n^{(Z_N)} = \frac{N\kappa_n^{(X_1)}}{N^{\frac{n}{2}}\sigma^n} \sim N^{1-\frac{n}{2}} \to 0$$

as $N \to \infty$

This is an indication that the Z variable acts like a normal distribution.

More precisely,

$$\ln \varphi_X(t) = it\mu + \frac{(it)^2}{2}\sigma^2 + \cancel{\quad}\,\psi(t)$$

$\psi(t) = o(t^2)$ something that dies faster than t^2 as $t \to 0$

$= t^2 \varepsilon(t)$

$$\varphi_{Z_N}(t) = E\left[e^{it\left(\frac{S_N - N\mu}{\sqrt{N}\sigma}\right)}\right]$$

$$= e^{-\frac{iNt\mu}{\sqrt{N}\sigma}} \varphi_{S_N}\left(\frac{it}{\sqrt{N}\sigma}\right)$$

$$= e^{-\frac{iNt\mu}{\sqrt{N}\sigma}} \left[\varphi_{X_1}\left(\frac{it}{\sqrt{N}\sigma}\right)\right]^N$$

$$= e^{-\frac{iNt\mu}{\sqrt{N}\sigma}} e^{N\left[\frac{it}{\sqrt{N}\sigma}\mu + \frac{(it)^2}{2N\sigma^2}\sigma^2 + \psi\left(\frac{t}{\sqrt{N}\sigma}\right)\right]}$$

$$= e^{-\frac{t^2}{2} + N\psi\left(\frac{t}{\sqrt{N}\sigma}\right)}$$

$$N\psi\left(\frac{t}{\sqrt{N}\sigma^2}\right) = N\frac{t^2}{N\sigma^2} \frac{\psi\left(\frac{t}{\sqrt{N}\sigma^2}\right)}{t^2/N\sigma^2}$$

$$= \frac{t^2}{\sigma^2} \underbrace{\frac{\psi(t/(\sqrt{N}\sigma))}{(t^2/N\sigma^2)}}_{\to 0 \text{ as } N\to\infty} = \frac{t^2}{\sigma^2}\xi\left(\frac{t}{\sqrt{N}\sigma}\right)$$

$$\lim_{x\to 0} \xi(x) = 0 \qquad x = \frac{t}{\sqrt{N}\sigma}$$

So $Z_N \sim \mathcal{N}(0,1)$ as $N \to \infty$

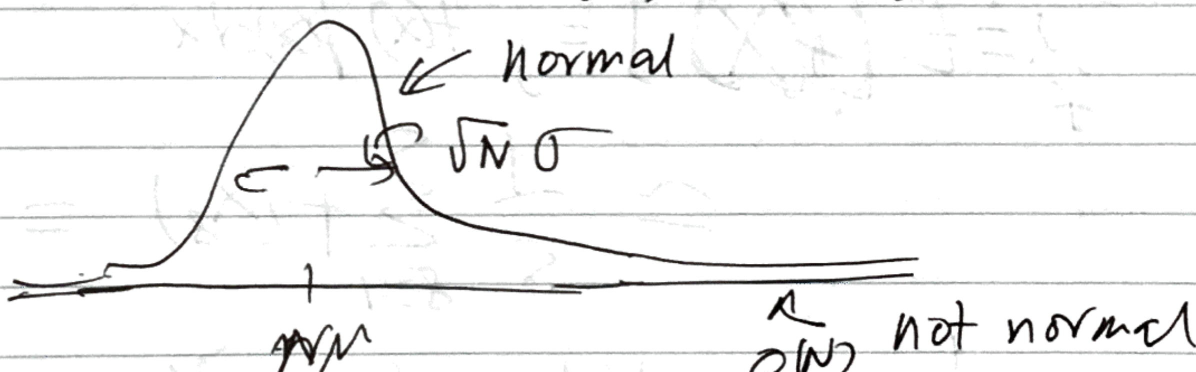
We already saw a special case: Binomial distribution

$$k = x_1 + \quad + x_n$$

indept. success/failure represented by 1/0.

$E[X_i] = \theta$ $\qquad$ $var[X_i] = \theta(1-\theta)$

$$Z_n = \frac{k - n\theta}{\sqrt{n\theta(1-\theta)}}$$ becomes normal as $n \to \infty$



By the way, tails outside the $\sqrt{N}$ size region not given by non-normal distribution. Large deviation theory!

Monte Carlo approximation

Set up a stochastic process / Markov chain, which has the stationary distribution which is your target prob. distrⁿ.

Then estimate

$$\mu_f = E[f(X)] = \int f(x)\, p(x)\, dx$$

$$\simeq \frac{1}{S} \sum_{s=1}^{S} f(x_s) = \hat{\mu}_f$$

by sampling from the Markov process.

If this samples are nearly ~~indep~~ <u>independent</u>,

$$\hat{\mu}_f \sim \mathcal{N}(\mu_f, \sigma_f^2/S)$$

Where $\sigma_f^2 = \mathrm{Var}[f(X)]$

$$\sigma_f^2 \approx \frac{1}{S}\sum_{s=1}^{S}\left(f(x_s)-\hat{\mu}_f\right)^2 = \hat{\sigma}_f^2$$

μ_f to $\hat{\mu}_f$ distance is order $\frac{\hat{\sigma}_f}{\sqrt{S}}$ which decreases as S increases

More precisely, one can make statements like.

$$P\left\{\hat{\mu}_f - 1.96\frac{\hat{\sigma}_f}{\sqrt{S}} \le \mu_f \le \hat{\mu}_f + 1.96\frac{\hat{\sigma}_f}{\sqrt{S}}\right\} \simeq 0.95$$

The samples being independent is important. For the example of measuring 'π', it is easy to draw random samples.

In case x's from the Markov chain are correlated in time, we often have to separate them in time for independence.