last change: oct.$13^{th}$ 2012@18.00

# The most important distributions of anamult

**Summary 1** *: binomial, hypergeometric, multinomial, Poisson, multiNormal, F, Chi2*

## hypergeometric

$$X \sim \mathrm{HypG}\left(N, K, n\right)$$

motivation: population (size $N$) divided into two subpopulations (sizes $K$ and $N-K$ ). random sample (size $n$). **random** variable $X$ : number of elements of type $K$ in sample

$$\mathbb{P}\left(x\right) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

alternative notation

$$\mathbb{P}\left(x_1, x_2\right) = \frac{\binom{K_1}{x_1}\binom{K_2}{x_2}}{\binom{K_1+K_2}{x_1+x_2}}$$
$$x_1 + x_2 = n$$
$$K_1 + K_2 = N$$

## multivariate hypergeometric

$$\mathbb{P}\left(x_1, \ldots, x_j\right) = \frac{\binom{K_1}{x_1}\binom{K_2}{x_2}\cdots\binom{K_j}{x_j}}{\binom{K_1+\cdots+K_j}{x_1+\ldots+x_j}}$$
$$x_1 + \ldots + x_j = x$$
$$K_1 + \cdots + K_j = K$$

## binomial distribution

$$X \sim \mathrm{Bnom}(p, n)$$

$\Longrightarrow$

$$\mathbb{P}(x) = \binom{n}{x} p^x (1-p)^{n-x}$$
$$\binom{n}{n_1} = \frac{n!}{n_1!(n-n_1!)}$$

alternative notation

$$\mathbb{P}(x_1, x_2) = \binom{n}{x_1, x_2} p_1^{x_1} p_2^{x_2}$$
$$p_1 + p_2 = 1$$
$$x_1 + x_2 = n$$

## multinomial distribution

$$\mathbf{X} \sim \mathrm{Mnom}_q\,(n,\ p_1, \ldots, p_q)$$

density

$$\mathbb{P}(x_1, ..., x_q) = \binom{n}{x_1, ..., x_q} p_1^{n_1} \cdots p_q^{n_q}$$
$$\sum_j p_j = 1$$
$$\sum_j x_j = n$$

with

$$\binom{n}{x_1, ..., x_q} = \frac{n}{x_1! ... x_q!} = \binom{n}{x_1}\binom{n - x_1}{x_2}\binom{n - x_1 - x_2}{x_3} ... \underbrace{\binom{n - x_1 ... - x_{q-1}}{x_q}}_{=\binom{x_q}{x_q}=1}$$

alternative notation:

$$\mathbb{P}(x_1, ..., x_q) = \binom{n}{x_1, ..., x_q} p_1^{n_1} \cdots p_{q-1}^{n_{q-1}} \left(1 - p_1^{n_1} - \cdots - p_{q-1}^{n_{q-1}}\right)$$

## Poisson distribution

$$X \sim \mathrm{Pois}(\lambda)$$

rate $\lambda$

$$\mathbb{P}(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

DeMoivre-Laplace: (law of rare events)

$$\left.\begin{matrix} n \to \infty \\ p \to 0 \\ np \to \lambda \end{matrix}\right\} \Longrightarrow \binom{n}{n_1} p^{n_1} (1-p)^{n-n_1} \approx \frac{\lambda^{n_1}}{n_1!} e^{-\lambda}$$

## Univariate Normal Distribution

density (standard normal)

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\}$$

general case: $y = \sigma x + \mu \Longrightarrow f(y) = f(x(y)) \, |dx/dy| \Longrightarrow$

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}\right\}$$

## Multivariate Normal Distribution

$$\mathbf{X} \sim \mathcal{N}_p(\mu, \Sigma) \implies f(\mathbf{x} \mid \mu, \Sigma) = \frac{1}{(2\pi \ \ |\Sigma|)^{p/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)' \Sigma^{-1} (\mathbf{x}-\mu)\right)$$

Special case: dimension $p = 2$

$$\begin{bmatrix} X \\ Y \end{bmatrix} \backsim \mathcal{N}_p\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}\right)$$

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

where

$$\begin{aligned} \mu_1 &= \mathbb{E}X, \ \ \mu_2 = \mathbb{E}Y \\ \sigma_1^2 &= \mathbb{V}X, \ \ \sigma_2^2 = \mathbb{V}X, \ \ \sigma_{12} = \mathbb{C}(X,Y) \\ \rho &= \frac{\sigma_{12}}{\sigma_1\sigma_2} \end{aligned}$$

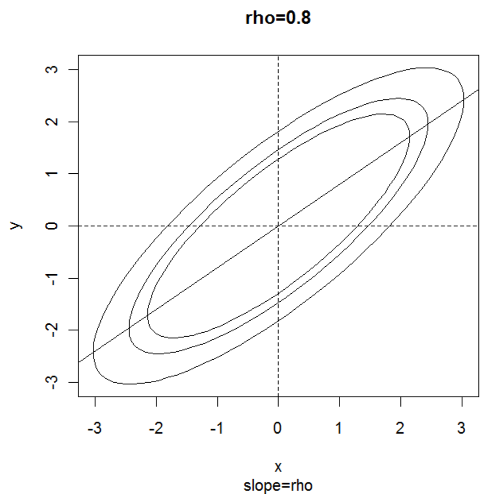
further distributions
$\chi^2-$distribution: "sum of squares of normals"

$$z_1, \ldots, z_p \text{ iid } \sim \mathcal{N}(0,1) \implies \chi^2 = \sum_i^p z_i^2 \sim \chi^2(p)$$

also

$$x_1, \ldots, x_p \text{ iid } \sim \mathcal{N}(\mu, \sigma^2) \implies \chi^2 = \sum_i^p \left(\frac{x_i - \mu}{\sigma}\right)^2 \sim \chi^2(p)$$

$F-$distribution: "ratio of sum of squares of normals"




```
library(ellipse)
rho<-0.8
plot(ellipse(rho,level=0.99),type="l")
```

Ellipse and Regression line

# Appendix: **Fisher test**.

Data

| | male | fmale | total |
|---|---|---|---|
| dieting | 1 | 9 | 10 |
| ndieting | 11 | 3 | 14 |
| totals | 12 | 12 | 24 |

draw sample of size 12 of population of size 24=12male+12fmale. find probability of $x$ dieting males if there are 24=10diet+14ndiet

Compute the likelihood of your data under the equally likely hypothesis.

**A)** $\binom{24}{12}$ choices to give label "male" to 12 out of 24
**B)** $\binom{10}{x}\binom{14}{12-x}$ choices to give label "diet" and "ndiet" to the chosen "male"

$$\mathbb{P}\,(\text{dieting males} = x) = \frac{\binom{10}{x}\binom{14}{12-x}}{\binom{24}{12}}$$

note: if dieting among males and females equally likely, expect half of dieters to be male and half of non-dieters to be male.

**CapRecap experiments**

*How many fish in a lake?*
**1)** catch $x$ and mark them, return them to the lake
**2)** catch $n$ and count the marked ones

**3)** denote the total number of fish by $N = x + (N - x)$
probabilty of catching $y$ marked fish

$$P(y) = \frac{\binom{x}{y}\binom{N-x}{n-y}}{\binom{N}{n}}$$

estimate $\widehat{N} = \arg\max\ P(y)$