Experimental Design and Statistical Analysis (EDSA)
The Graduate School of CAAS

Lecture 2

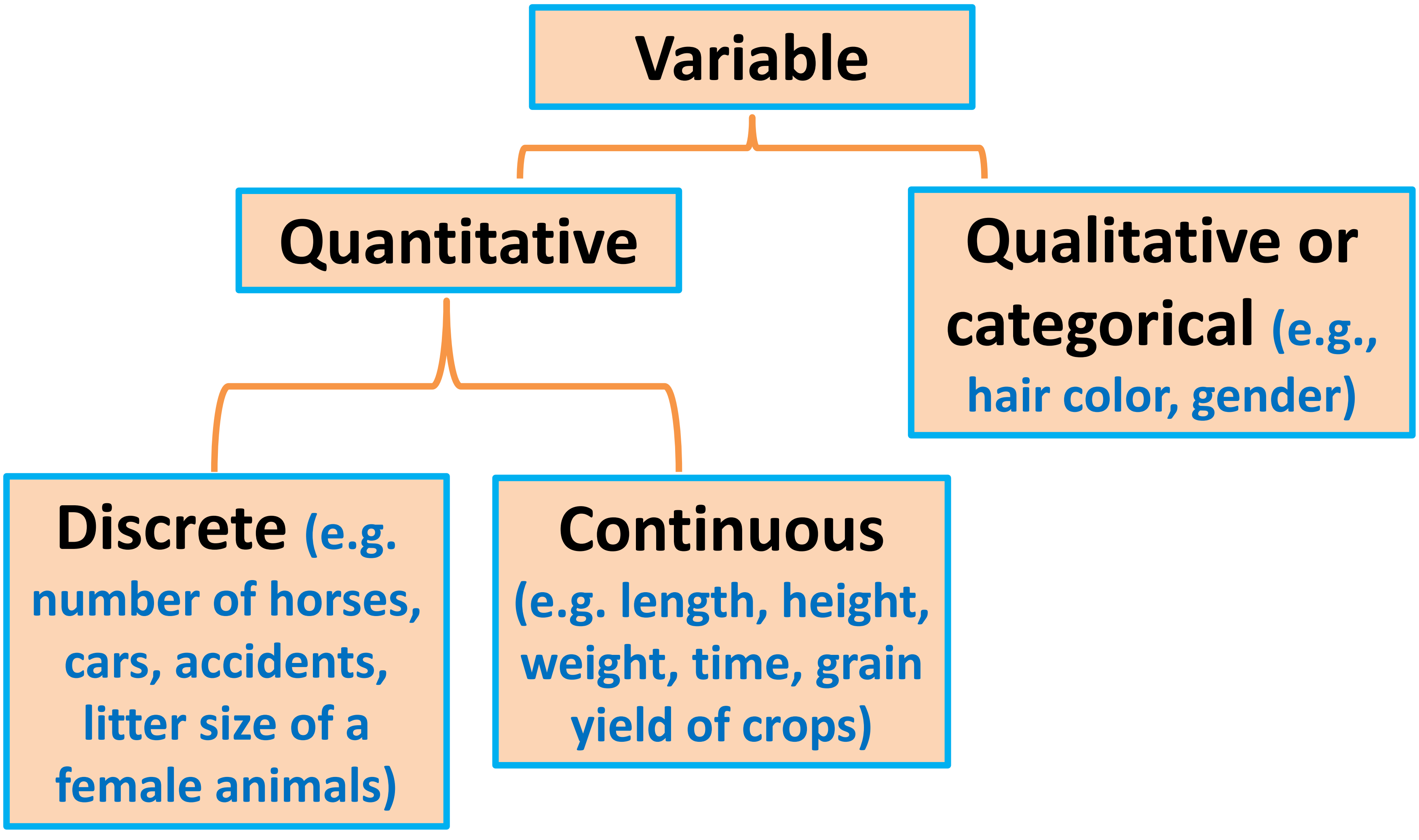
Population and its distribution

Variables

Variables and constants

- A **variable** is a characteristic under study that assumes different values for different elements.
- In contrast to a variable, the value of a constant is fixed, such as planting date
- The value of a variable for an element is called an observation or measurement.
- A data set is a collection of observations on one or more variables.

Types of variables



Variables

- The **independent variable** is the variable that is purposely changed. It is the manipulated variable. E.g. amount of fertilizer, planting date and density, genotype etc.
- The **dependent variable** changes in response to the independent variable. It is the responding variable. E.g. grain yield and quality.

Single Variable vs. Multiple Variables

- Single Variable:
  - Only one independent variable
  - Cannot look at interactions
- Multiple Variables:
  - Two or more independent variables
  - If use factorial design, can look at interactions
  - Can require a lot of participants (between) or time (within)

Independence of data

- Samples tell us about Populations
- This is only true if the data in a sample are drawn randomly from the population
- The true difficulty of non-independent data is that we do not know how it influences the sample (could be positively or negatively correlated)

Interactions

- Two independent variables **interact** when the effect of one depends on the level of the other

Basic statistics

- Random phenomena and random variables
- Two types of random variables
- Discrete variable
- Some common distributions of discrete variable
- Continuous random variable
- Some common distributions of continuous variable
- Distribution function
- Mean
- Variance and Covariance
- Correlation coefficient

Random phenomena and random variables

- We can not predict the result of random phenomenon
  - For example: flipping a coin, before they fall on the table, we do not know it is top or head.
- Random variable is used to denote random phenomenon
  - For example: set ***r.v.*** X as the result of flipping a coin
  - X=1 for top
  - X=0 for head

Two types of random variables

- Discrete random variable
- Continuous random variable

Each trait on the right is controlled by one pair of genes.

The seven loci are independent in genetics, i.e. no linkage between them.

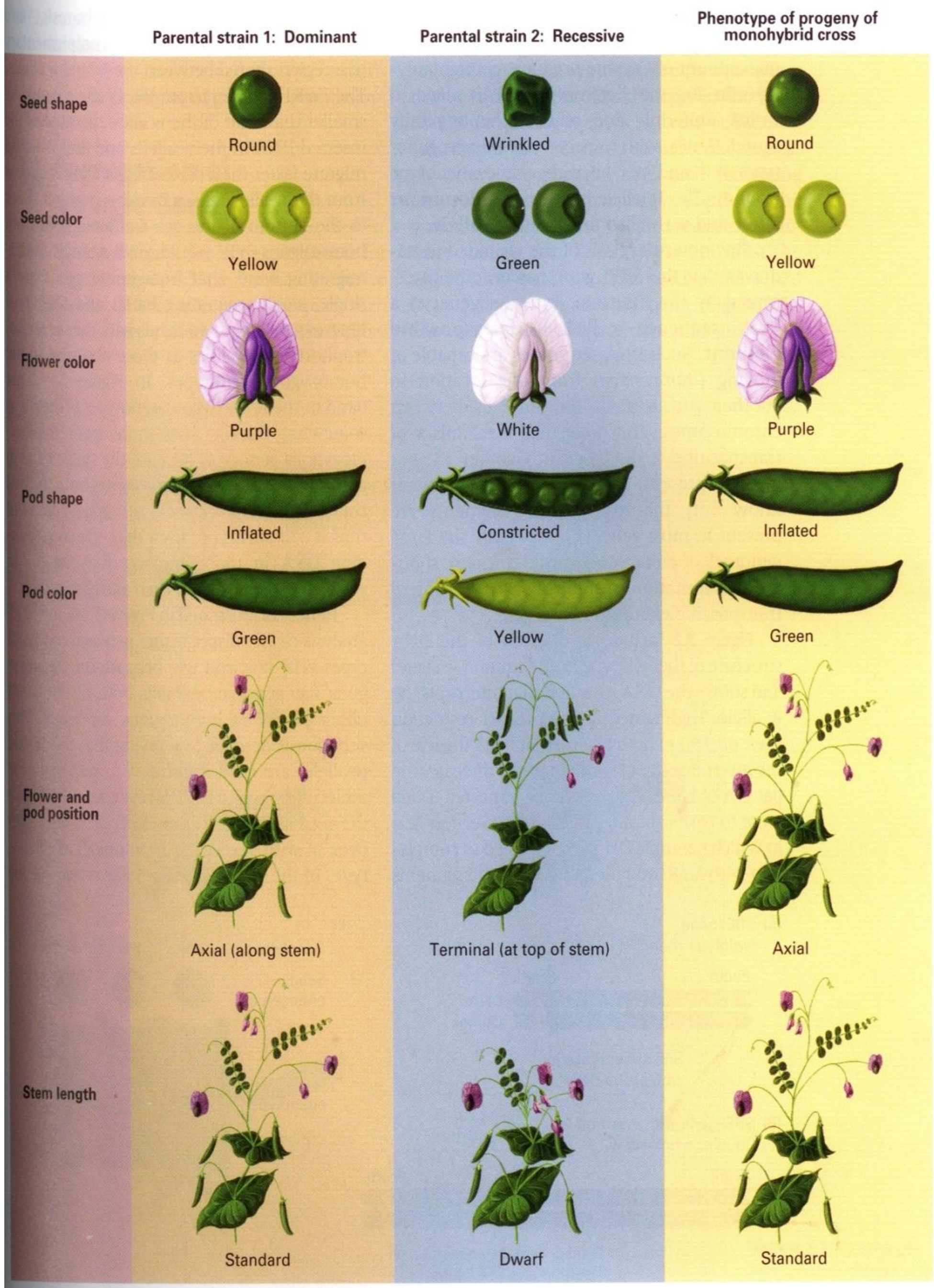
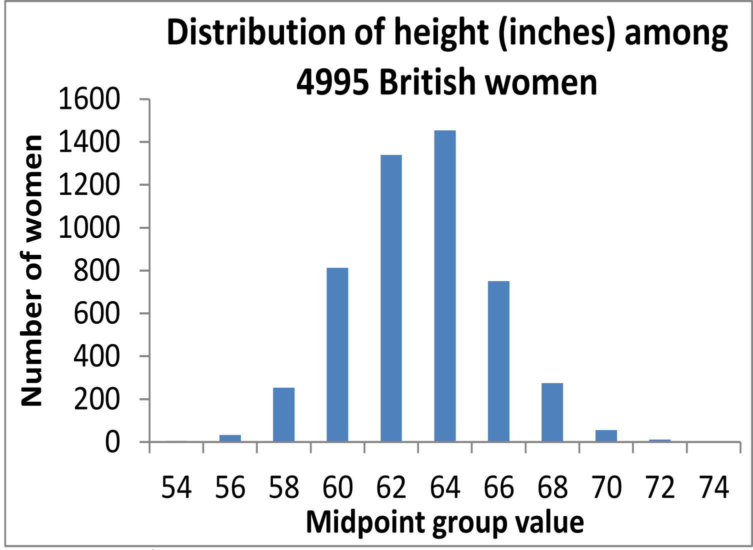
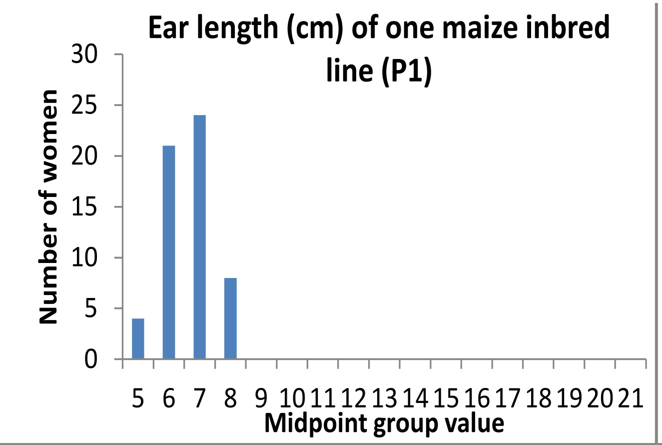
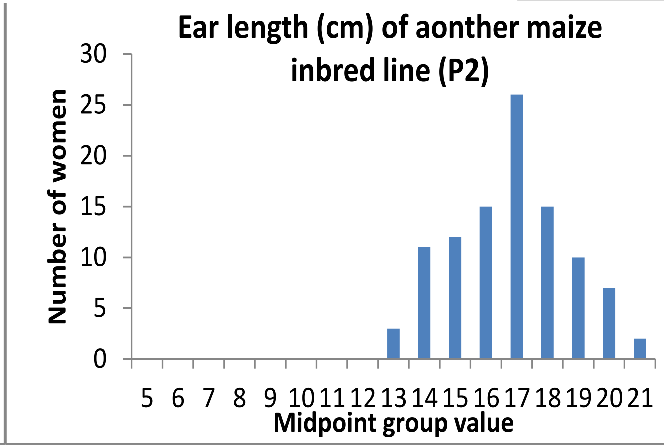


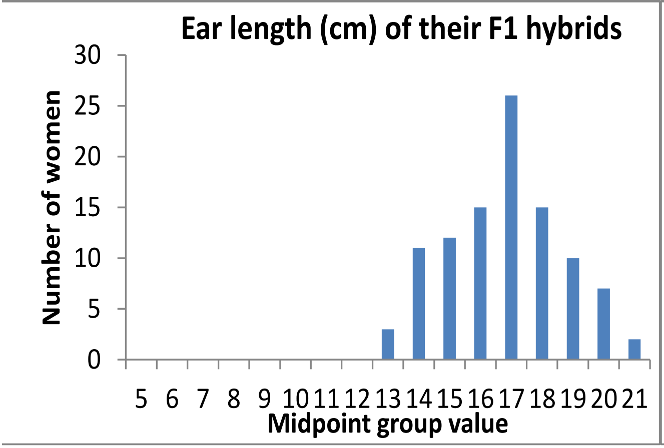
Figure 3.1 The seven different traits in peas studied by Mendel. The phenotype shown at the far right is the dominant trait, which appears in the hybrid produced by crossing.

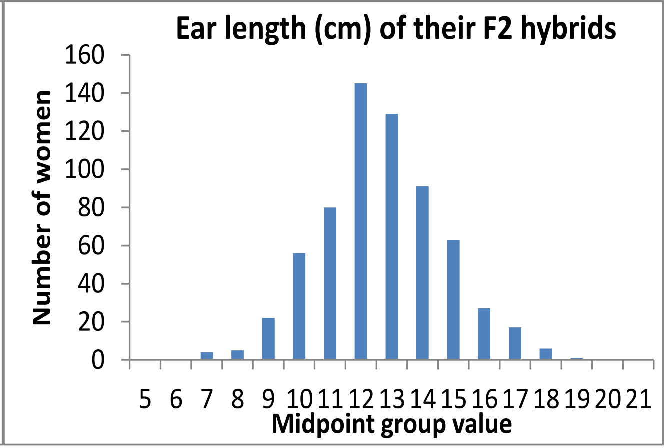
Quantitative traits











Discrete random variable

- Discrete random variable X that takes on a countable number of values $x_1, x_2, \ldots, x_k, \ldots$
- We wonder the probability of each value

$(k = 1, 2, \ldots)$

- Probability Distribution of X

$$P(X = x_k) = p_k$$

X	x_1	x_2	$\ldots$	x_k	$\ldots$
$\boldsymbol{P}$	p_1	p_2	$\ldots$	p_k	$\ldots$

Two basic properties of discrete variable

- (1) $p_k \geq 0$, k = 1, 2, …, indicating that a probability cannot be negative!
- (2) $\sum_{k=1}^{\infty} p_k = 1$, indicating sum of the probabilities of all possible values is 1!

Two-point distribution, or Bernoulli distribution

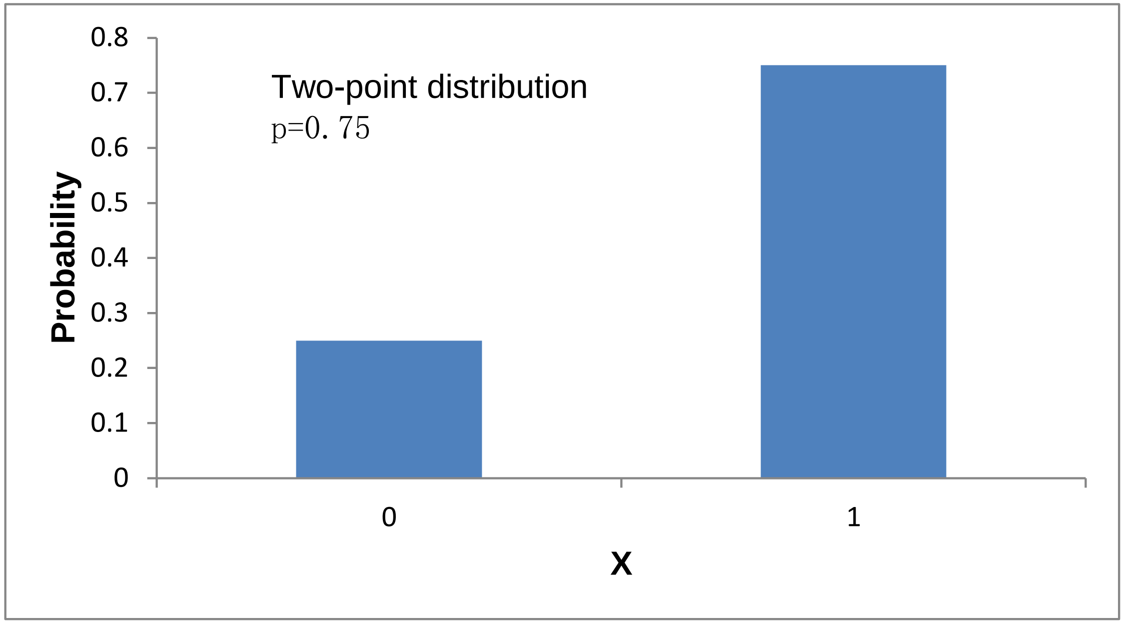
- The random variable has only two values: 0 and 1.

$$P(X=1)=p, P(X=0)=q, 0<p<1, p+q=1$$

$$\text{or } P(X=k)=p^k q^{1-k}, k=0,1$$

- In table:

X	0	1
P	q	p



Binomial Distribution

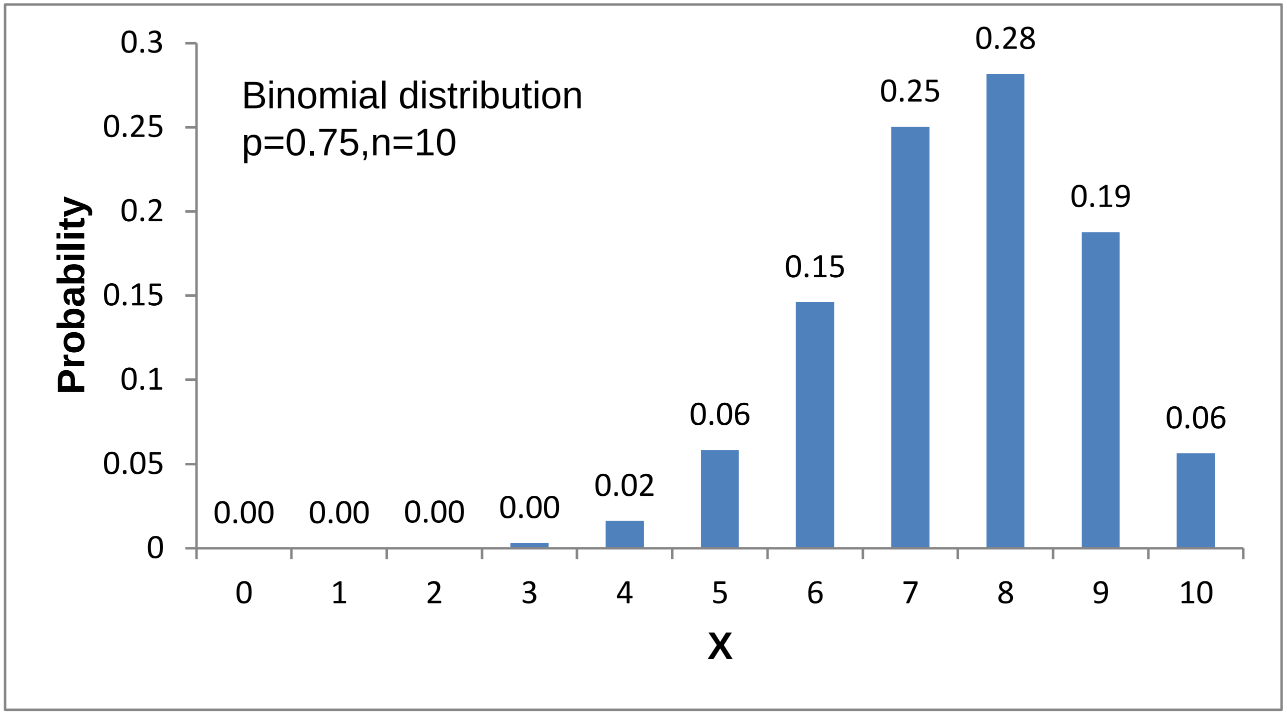
- The binomial distribution is the probability distribution that results from doing a "binomial experiment". Binomial experiments have the following properties:
  - Fixed number of trials, represented as n.
  - Each trial has two possible outcomes, a "success" and a "failure".
  - $P(\text{success})=p$ (and thus: $P(\text{failure})=1-p$), for all trials.
  - The trials are independent, which means that the outcome of one trial does not affect the outcomes of any other trials.

Binomial Distribution

- If the distribution of variable X satisfies

$$P(X=k)=C_n^k p^k q^{n-k}, k=0,1,\cdots,n; 0<p<1, q=1-p$$

- Then X follows the Binomial distribution.



Poisson Distribution

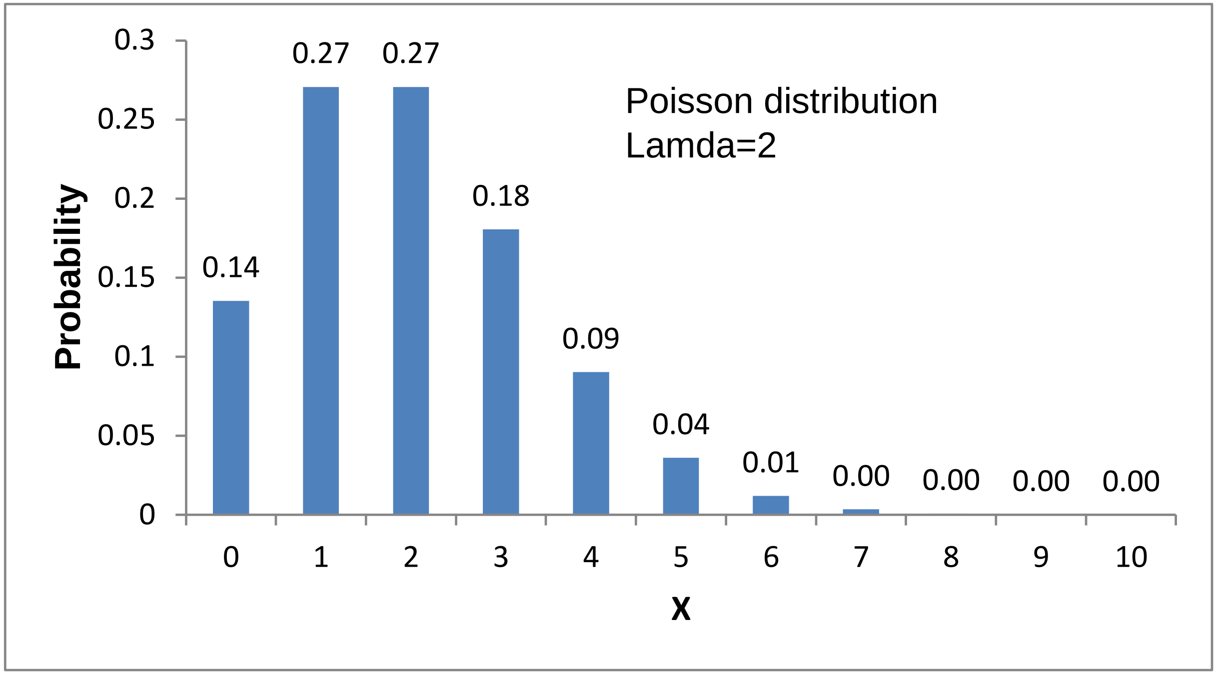
- Named for Simeon Poisson, the Poisson distribution is a discrete probability distribution and refers to the number of events within a specific time period or region of space. For example:
  - The number of cars arriving at a service station in 1 hour. (The interval of time is 1 hour.)
  - The number of accidents in 1 day on a particular stretch of highway. (The interval is defined by both time, 1 day, and space, the particular stretch of highway.)

Poisson Distribution

If the distribution of variable X satisfies

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \lambda>0,\ k=0, 1, 2, \ldots$$

then X follows the Poisson distribution.



Continuous variable

- If a variable can take on any value between two specified values, it is called a continuous variable.
- For a random variable X, if there is a non-negative Lebesgue-integrable function

$$p(x) \quad -\infty < x < +\infty$$

for any a<b, we have

$$P(a < x < b) = \int_a^b p(x)dx$$

then we say X is a continuous variable, and we call $p(x)$ the probability density function (p.d.f.) of X

Three basic properties of continuous variable

- (1) $p(x) \geq 0$, indicating that a probability density cannot be negative!
- (2) $\int_{-\infty}^{+\infty} p(x)dx = 1$, indicating the integral of the probability density of all possible values is 1!
- (3) For any a, $P(x = a) = \int_{a}^{a} p(x)dx = 0$

So for any $x_1 < x_2$,

$P(x_1 < x < x_2)$ = $P(x_1 \leq x < x_2)$ = $P(x_1 < x \leq x_2)$ = $P(x_1 \leq x \leq x_2)$

Uniform distribution

- For a random variable X, if its probability density satisfies

$$p(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & others \end{cases}$$

then X obeys the Uniform distribution on [*a*, *b*].

Exponential Distribution

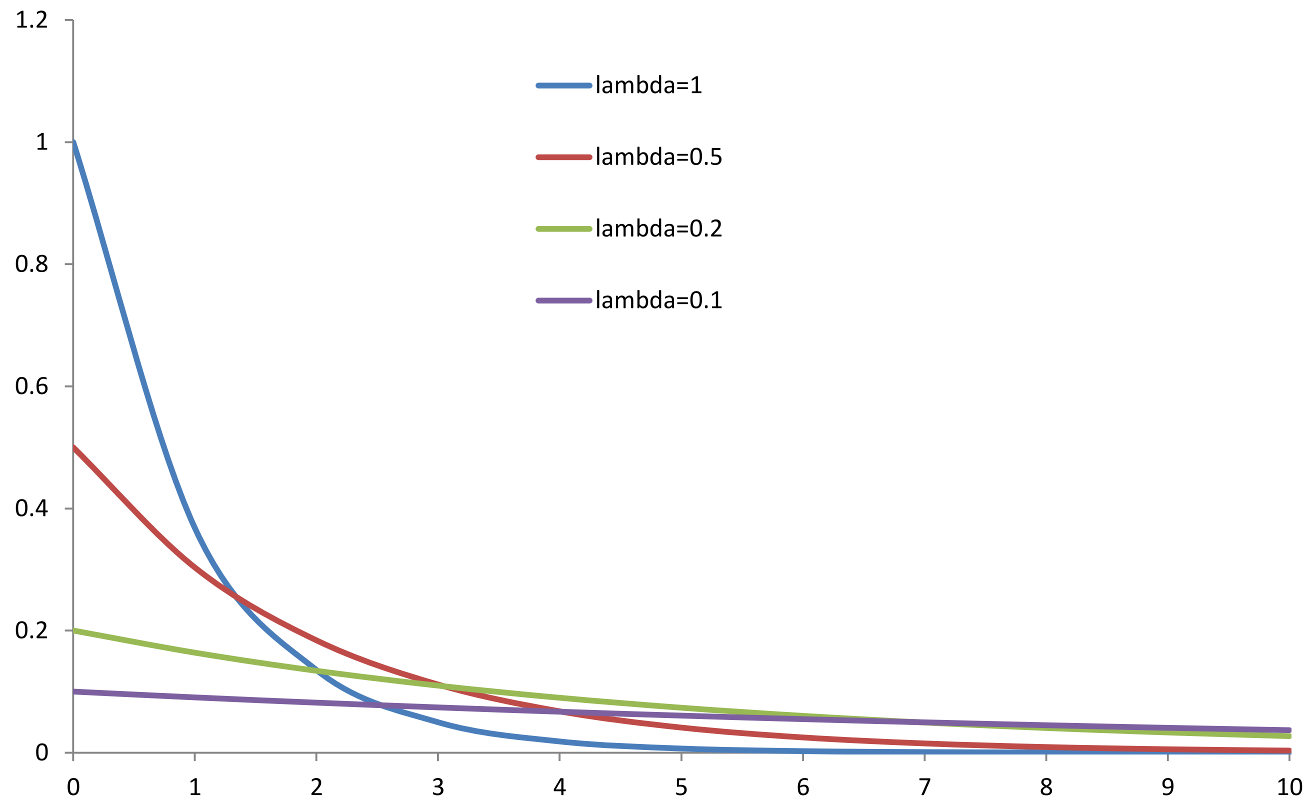
- Random variable can only take on positive values, e.g. life expectation of a bulb light, a person etc.
- Used to model inter-arrival times/distances for a Poisson process
- Its probability density satisfies

$$p(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

Exponential Density Functions (p.d.f.)

- Lack-of-memory of Exponential distribution

$$p(x > s + t \mid x > s) = \frac{p(x > s + t)}{p(x > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t}$$

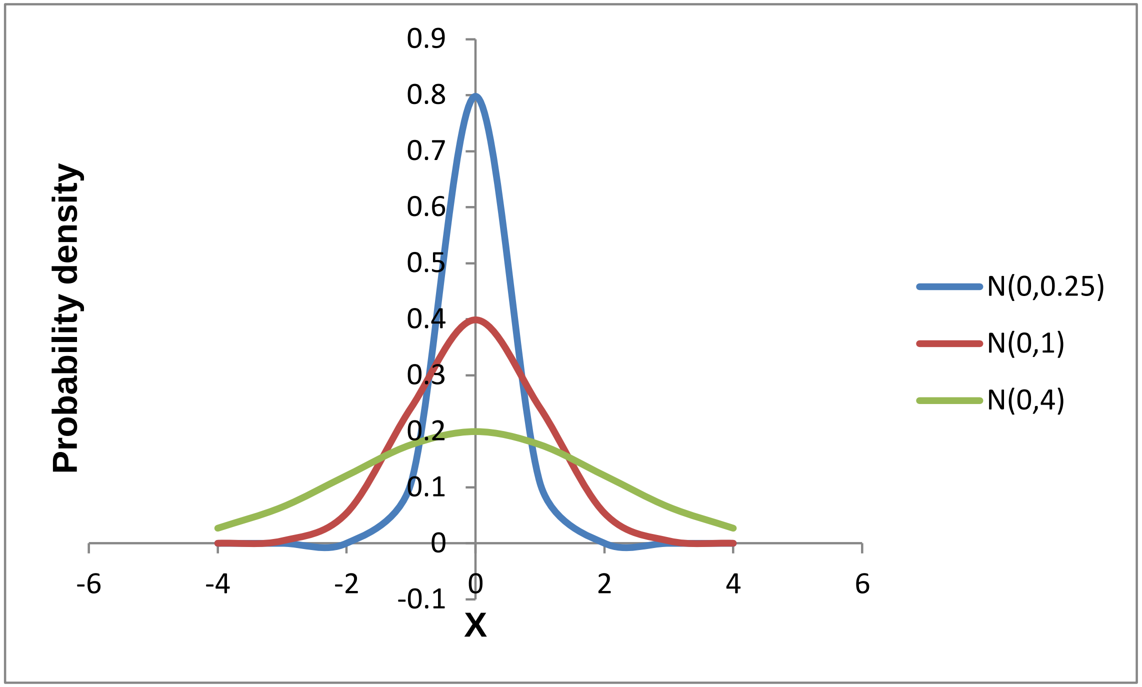


Normal distribution

- For a random variable X, if its probability density satisfies

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < +\infty, \quad \sigma^2 > 0$$

then X obeys the normal distribution X~ $N(\mu, \sigma^2)$



Standardized normal distribution

- Standard normal distribution is a normal distribution with mean 0 and standard deviation 1. i.e. $N(0,1)$
- The probability density satisfies $p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
- The probabilities of X in some intervals are:

Interval	Probability
[-1, 1]	0.6827
[-2, 2]	0.9545
[-1.96, 1.96]	0.95
[-2.58, 2.58]	0.99
[-3, 3]	0.9973

Gamma Function

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

$$\Gamma(\alpha+1) = \int_0^\infty x^{\alpha} e^{-x} dx \quad \text{Integrating by Parts:}$$

$$u = x^{\alpha} \Rightarrow du = \alpha x^{\alpha-1} dx$$

$$dv = e^{-x} dx \Rightarrow v = -e^{-x}$$

$$\Rightarrow \Gamma(\alpha+1) = \int_0^\infty x^{\alpha} e^{-y} dx = uv - \int v du = -x^{\alpha} e^{-x}\Big|_0^\infty + \int_0^\infty \alpha x^{\alpha-1} e^{-x} dx =$$

$$= -0 - (-0) + \alpha \int_0^\infty x^{\alpha-1} e^{-x} dx = \alpha \Gamma(\alpha) \ \text{(Recursive Property)}$$

Note that if α is an integer, $\Gamma(\alpha) = (\alpha - 1)!$

Consider the integral: $\int_0^\infty x^{\alpha-1} e^{-x/\beta} dx$ Letting $y = x/\beta$:

$$\Rightarrow x = y\beta \Rightarrow dx = \beta dy$$

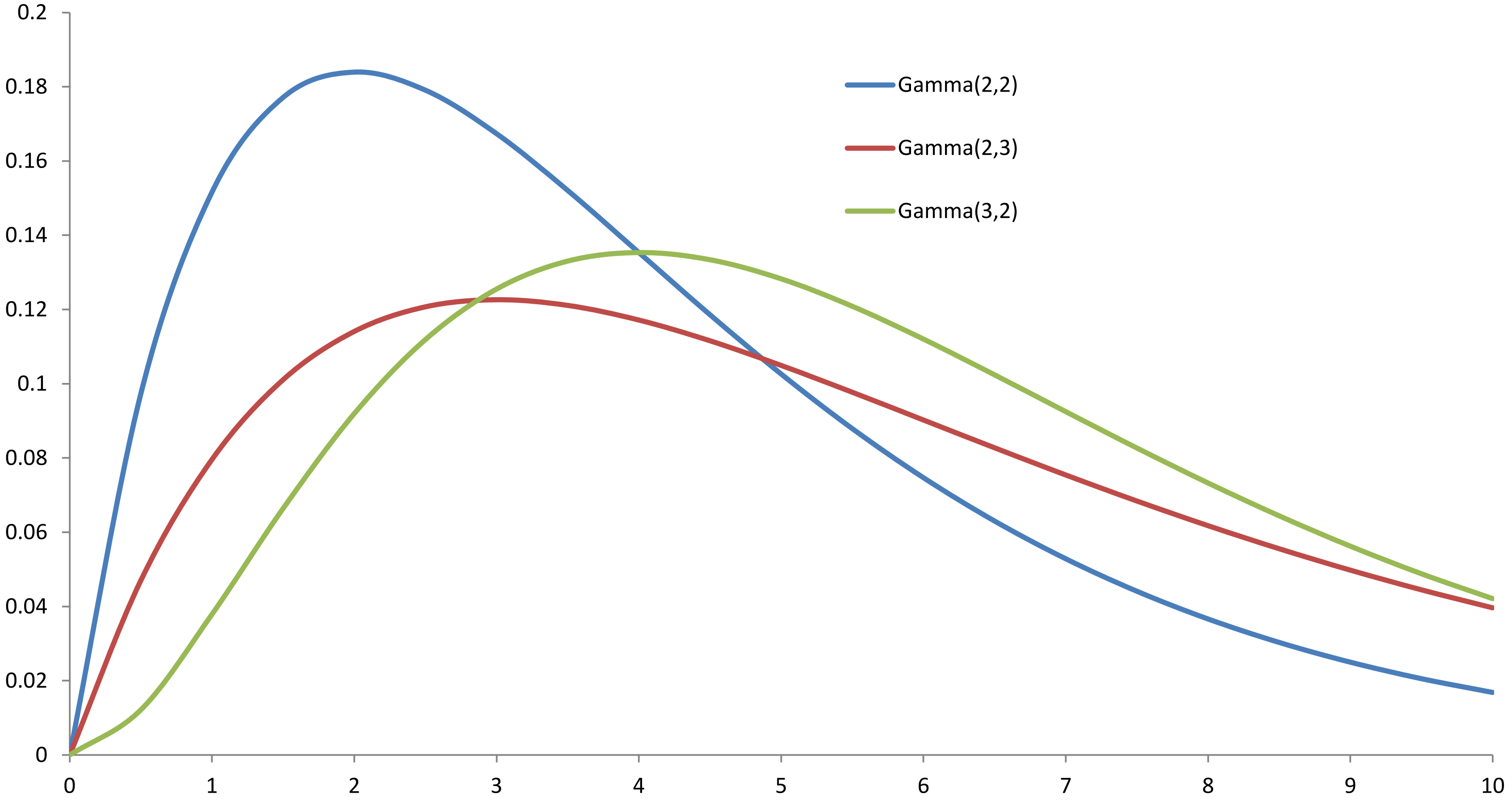
$$\Rightarrow \int_0^\infty x^{\alpha-1} e^{-x/\beta} dx = \int_0^\infty (y\beta)^{\alpha-1} e^{-y} \beta dy = \beta^{\alpha} \int_0^\infty y^{\alpha-1} e^{-y} dy = \beta^{\alpha} \Gamma(\alpha)$$

Gamma Distribution

- Random Variable can take on positive values only
- Used to model many biological and economic characteristics
- Can take on many different shapes to match empirical data
- Its probability density satisfies

$$p(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} & x > 0; \alpha, \beta > 0 \\ 0 & \text{others} \end{cases}$$

Gamma Densities (p.d.f.)



Beta distribution

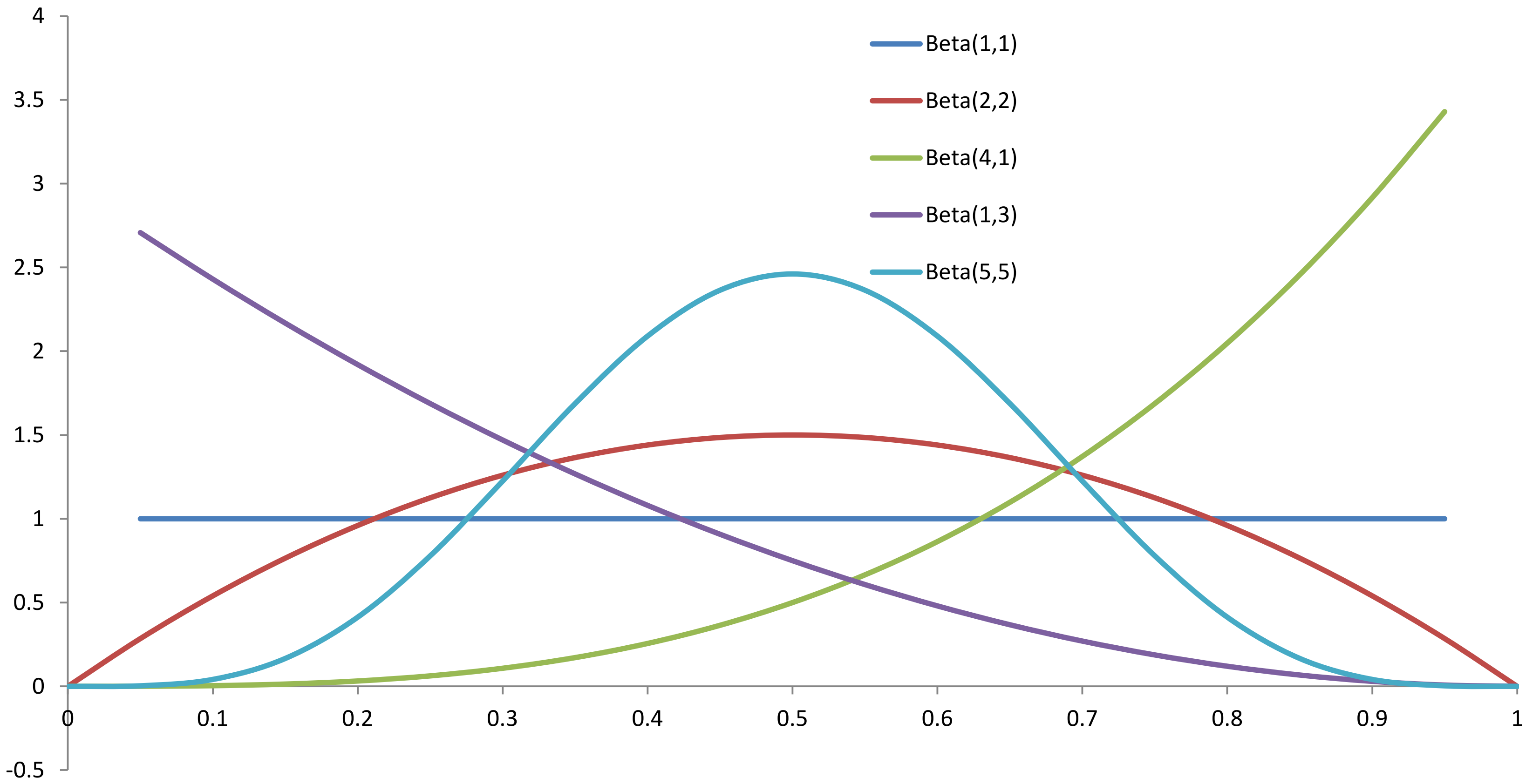
- Used to model probabilities (can be generalized to any finite, positive range)
- Parameters allow a wide range of shapes to model empirical data

$$p(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, 0 < x < 1; \alpha, \beta > 0 \\ 0 \qquad \text{otherwise} \end{cases}$$

$$\text{Denote } B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx,$$

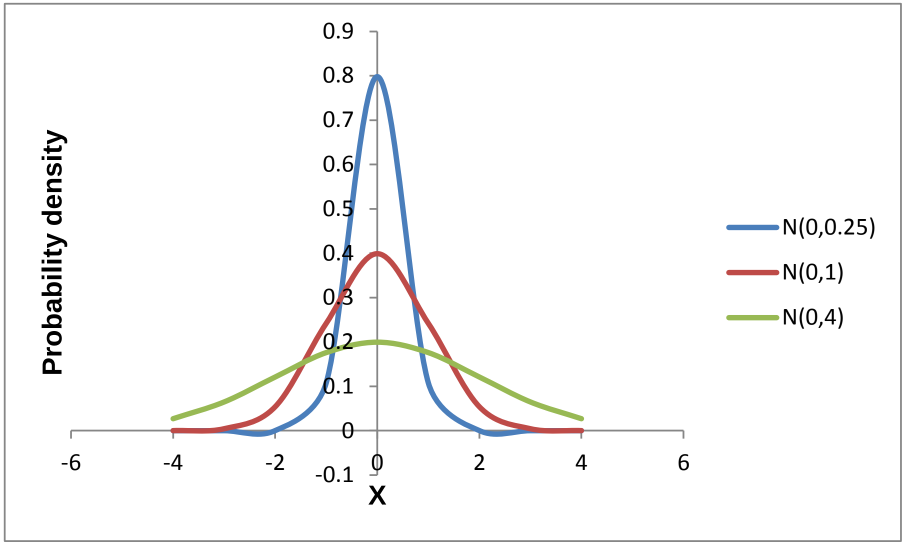
$$\text{we can prove that } B(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

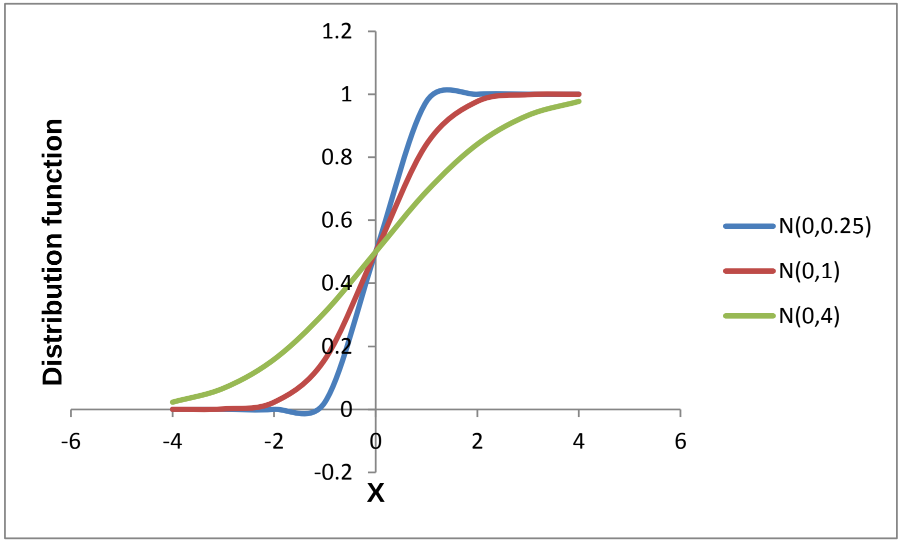
Beta Density Functions (p.d.f.)



Distribution function

- X is a continuous variable with probability density $p(x)$, $-\infty < x < \infty$, then its distribution function is $F(x) = P(X \le x) = \int_{-\infty}^{x} p(t)dt$





Mean of a distribution

- Mean, which is also called expectation, provides the average of the variable.
- If X is discrete,

$$E(X) = \sum_{k=1}^{\infty} x_k p_k$$

- If X is continuous,

$$E(X) = \int_{-\infty}^{+\infty} xp(x)dx$$

Properties of the mean

- 1. For any constant C, $E(C)=C$
- 2. X is a random variable, C is a constant, then $E(CX)=CE(X)$
- 3. X and Y are random variables, then $E(X+Y) = E(X) + E(Y)$
- 4. X and Y are independent, then $E(XY) = E(X) \times E(Y)$

Variance and covariance

- The variance is a measure of how far a set of numbers is spread out.

$$V(X) = E[X - E(X)]^2 = E(X^2) - [E(X)]^2$$

- The covariance between random variables X and Y is

$$Cov(X,Y) = E\{[X - E(X)] \times [Y - E(Y)]\} = E(XY) - E(X)E(Y)$$

- If *Cov*(X, Y)=0, then X and Y are called to be independent. Or there is no linear relationship between X and Y.

Properties of variance

- 1. For any constant C, $V(C)=0$
- 2. X is a random variable, C is a constant, then $V(CX)=C^2V(X)$
- 3. X and Y are independent, then $V(X+Y)=V(X)+V(Y)$
- 4. X and Y are random variable, then $V(X+Y)=V(X)+V(Y)+2Cov(X, Y)$
- Notes: $V(XY)\neq V(X)V(Y)$, even if X and Y are independent

Properties of covariance

- 1. $Cov(X, Y) = Cov(Y, X)$
- 2. $Cov(aX, bY) = abCov(X, Y)$
- 3. $Cov(X+Y, Z) = Cov(X, Z)+Cov(Y, Z)$

Correlation coefficient

- Correlation coefficient (Pearson's correlation coefficient) is defined as the covariance of the two variables divided by the product of their standard deviations:

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E\left[(X-\mu_X)(Y-\mu_Y)\right]}{\sigma_X \sigma_Y}$$

$$\sigma_X = \sqrt{V(X)}, \sigma_Y = \sqrt{V(Y)}$$

Means and variances of some common distributions

Bernoulli (or two point) distribution

$$\mathrm{P}(X = k) = p^k q^{1-k} \; (q = 1 - p, k = 0, 1)$$

- So

$$E(X) = \sum_{k=0}^{1} k p^k q^{1-k} = p$$

$$E(X^2) = \sum_{k=0}^{1} k^2 p^k q^{1-k} = p$$

$$V(X) = p - p^2 = p(1 - p)$$

Binomial distribution

$$\mathrm{P}(\mathrm{X}=k)=C_n^k p^k q^{n-k} \ (q=1-p, k=0,1,\cdots,n)$$

$$E(X)=\sum_{k=0}^{n} k C_n^k p^k q^{n-k} = \sum_{k=0}^{n} \frac{kn!}{k!(n-k)!} p^k q^{n-k}$$

$$= \sum_{k=0}^{n} \frac{n!}{(k-1)!(n-k)!} p^k q^{n-k}$$

$$= \sum_{k=0}^{n} \frac{np(n-1)!\, p^{k-1} q^{(n-1)-(k-1)}}{(k-1)![(n-1)-(k-1)]!}$$

$$\underline{\underline{\text{Let } k'=k-1}}\ np \sum_{k'=0}^{n-1} \frac{(n-1)!\, p^{k'} q^{(n-1)-k'}}{k'![(n-1)-k']!}$$

$$= np(p+q)^{n-1} = np$$

Binomial distribution (continued)

$$E(X^2) = \sum_{k=0}^{n} k^2 C_n^k p^k q^{n-k} = \sum_{k=0}^{n} k^2 \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

$$= \sum_{k=0}^{n} [(k-1)+1] \frac{n!}{(k-1)!(n-k)!} p^k q^{n-k}$$

$$= \sum_{k=0}^{n} (k-1) \frac{n!}{(k-1)!(n-k)!} p^k q^{(n\text{-}2)-(k\text{-}2)} + \sum_{k=0}^{n} \frac{n!}{(k-1)!(n-k)!} p^k q^{n-k}$$

$$\underline{\underline{k' = k-2}}\; n(n-1)p^2 \sum_{k'=0}^{n\text{-}2} \frac{(n-2)!}{k'!(n-2\text{-}k')!} p^{k'} q^{n-2-k'} + np$$

$$= n(n-1)p^2 (p+\mathrm{q})^{n\text{-}2} + np$$

$$= n(n-1)p^2 + np$$

$$V(X) = n^2p^2 - np^2 + np - n^2p^2 = np(1-p) = npq$$

Poisson distribution

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$$

- So

$$EX=\sum_{k=0}^{\infty}k\frac{\lambda^k}{k!}e^{-\lambda}=\sum_{k=1}^{\infty}k\frac{\lambda^k}{k!}e^{-\lambda}=\lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}$$

$$\overset{j=k-1}{=}\lambda e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^j}{j!}=\lambda e^{-\lambda}e^{\lambda}=\lambda$$

$$EX^2=\sum_{k=0}^{\infty}k^2\frac{\lambda^k}{k!}e^{-\lambda}=\sum_{k=1}^{\infty}k\frac{\lambda^k}{(k-1)!}e^{-\lambda}$$

$$=\sum_{k=1}^{\infty}(k-1+1)\frac{\lambda^k}{(k-1)!}e^{-\lambda}=\sum_{k=2}^{\infty}\frac{\lambda^k}{(k-2)!}e^{-\lambda}+\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}e^{-\lambda}$$

$$\overset{j=k-2}{=}\lambda^2e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^j}{j!}+\lambda=\lambda^2+\lambda$$

$$V(X)=\lambda^2+\lambda-\lambda^2=\lambda$$

Uniform Distribution

$$E(X) = \int_a^b x\left(\frac{1}{b-a}\right)dx = \left(\frac{1}{b-a}\right)\frac{x^2}{2}\Bigg|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{b+a}{2}$$

$$E\left(X^2\right) = \int_a^b x^2\left(\frac{1}{b-a}\right)dx = \left(\frac{1}{b-a}\right)\frac{x^3}{3}\Bigg|_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{(b-a)(a^2+b^2+ab)}{3(b-a)} =$$

$$= \frac{(a^2+b^2+ab)}{3}$$

$$\Rightarrow V(X) = E\left(X^2\right) - [E(X)]^2 = \frac{(a^2+b^2+ab)}{3} - \left[\frac{b+a}{2}\right]^2 =$$

$$= \frac{4(a^2+b^2+ab)-3(b^2+a^2+2ab)}{12} = \frac{a^2+b^2-2ab}{12} = \frac{(b-a)^2}{12}$$

Exponential Distribution

$$E(X) = \int_0^\infty x\left(\lambda e^{-\lambda x}\right)dx = \lambda \int_0^\infty x e^{-\lambda x} dx = \int_0^\infty x^{2-1} e^{-\lambda x} dx =$$

$$= \lambda \Gamma(2) \lambda^{-2} = \lambda^{-1}(2-1)! = \frac{1}{\lambda}$$

$$E\left(X^2\right) = \int_0^\infty x^2\left(\lambda e^{-\lambda x}\right)dx = \lambda \int_0^\infty x^2 e^{-\lambda x} dx = \int_0^\infty x^{3-1} e^{-\lambda x} dx =$$

$$= \lambda \Gamma(3) \lambda^{-3} = \lambda^{-2}(3-1)! = \frac{2}{\lambda^2}$$

$$\Rightarrow V(X) = E\left(X^2\right) - \left[E(X)\right]^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

where $\Gamma(\alpha) = \int_0^\infty \mathrm{x}^{\alpha-1} e^{-\mathrm{x}} dx$

Normal distribution

$$X \sim N(\mu, \sigma^2)$$

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

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

$$\underline{\underline{x-\mu = t}} \ \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (t+\mu) \exp\{-\frac{t^2}{2\sigma^2}\} \mathrm{dt}$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} t \exp\{-\frac{t^2}{2\sigma^2}\} dt + \frac{\mu}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\{-\frac{t^2}{2\sigma^2}\} dt$$

$$= -\frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} d \exp\{-\frac{t^2}{2\sigma^2}\} + \mu = 0 + \mu = \mu$$

Normal distribution (continued)

$$V(X) = E[X - \mathrm{E}(X)]^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 p(x) dx$$

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

$$\underline{\underline{t = \frac{x - \mu}{\sigma}}} \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t^2 \exp\{-t^2\} dt$$

$$= -\frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t d\exp\{-\frac{t^2}{2}\}$$

$$= -\frac{\sigma^2}{\sqrt{2\pi}} [(t\exp\{-\frac{\mathrm{t}^2}{2}\} |_{-\infty}^{+\infty}) - \int_{-\infty}^{+\infty} \exp\{-\frac{t^2}{2}\} dt]$$

$$= \sigma^2 \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{t^2}{2}\} dt = \sigma^2$$

Gamma distribution

$$E(X) = \int_0^\infty x\left(\frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}\right) dx = \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^\alpha e^{-x/\beta} dx =$$

$$= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{(\alpha+1)-1} e^{-x/\beta} dx = \frac{1}{\Gamma(\alpha)\beta^\alpha}\Gamma(\alpha+1)\beta^{\alpha+1} = \frac{\Gamma(\alpha+1)\beta}{\Gamma(\alpha)} =$$

$$= \frac{\alpha\Gamma(\alpha)\beta}{\Gamma(\alpha)} = \alpha\beta$$

$$E\left(X^2\right) = \int_0^\infty x^2\left(\frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}\right) dx = \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{\alpha+1} e^{-x/\beta} dx =$$

$$= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{(\alpha+2)-1} e^{-x/\beta} dx = \frac{1}{\Gamma(\alpha)\beta^\alpha}\Gamma(\alpha+2)\beta^{\alpha+2} = \frac{\Gamma(\alpha+2)\beta^2}{\Gamma(\alpha)} =$$

$$= \frac{(\alpha+1)\Gamma(\alpha+1)\beta^2}{\Gamma(\alpha)} = \frac{(\alpha+1)\alpha\Gamma(\alpha)\beta^2}{\Gamma(\alpha)} = (\alpha+1)\alpha\beta$$

$$\Rightarrow V(X) = E\left(X^2\right) - \left[E(X)\right]^2 = (\alpha+1)\alpha\beta - (\alpha\beta)^2 = \alpha^2\beta^2 + \alpha\beta^2 - \alpha^2\beta^2 = \alpha\beta^2$$

Beta distribution

$$E(X) = \int_0^1 \frac{x}{B(\alpha, \beta)} x^{\alpha-1}(1 - x)^{\beta-1} dx = \frac{1}{B(\alpha, \beta)} \int_0^1 x^{\alpha}(1 - x)^{\beta-1} dx$$

$$= \frac{B(\alpha + 1, \beta)}{B(\alpha, \beta)} = \frac{\Gamma(\alpha + 1)\Gamma(\beta)}{\Gamma(\alpha + \beta + 1)} \bigg/ \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

$$= \frac{\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha + \beta)\Gamma(\alpha + \beta)} \bigg/ \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} = \frac{\alpha}{\alpha + \beta}$$

$$E\left(X^2\right) = \int_0^1 \frac{x^2}{B(\alpha, \beta)} x^{\alpha-1}(1 - x)^{\beta-1} dx = \frac{1}{B(\alpha, \beta)} \int_0^1 x^{\alpha+1}(1 - x)^{\beta-1} dx$$

$$= \frac{B(\alpha + 2, \beta)}{B(\alpha, \beta)} = \frac{\Gamma(\alpha + 2)\Gamma(\beta)}{\Gamma(\alpha + \beta + 2)} \bigg/ \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

$$= \frac{\alpha(\alpha + 1)\Gamma(\alpha)\Gamma(\beta)}{(\alpha + \beta)(\alpha + \beta + 1)\Gamma(\alpha + \beta)} \bigg/ \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} = \frac{\alpha(\alpha + 1)}{(\alpha + \beta)(\alpha + \beta + 1)}$$

$$\Rightarrow V(X) = E\left(X^2\right) - \left[E(X)\right]^2 = \frac{\alpha(\alpha + 1)}{(\alpha + \beta)(\alpha + \beta + 1)} - \left(\frac{\alpha}{\alpha + \beta}\right)^2$$

$$= \frac{\alpha(\alpha + 1)(\alpha + \beta) - \alpha^2((\alpha + \beta + 1))}{(\alpha + \beta)^2(\alpha + \beta + 1)} = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Means and variances for some discrete distributions

Distribution	Probability function (p.d.)	Mean	Variance	Range of parameters
Two-point	P(X=x)= $p^x q^{1-x}$ x=0, 1	p	pq	q=1-p $0 < p < 1$
Binomial	$P(X=x)$= $C_n^x p^x q^{n-x}$ x=0, 1, 2, …, n	np	npq	q=1-p $0 < p < 1$
Poisson	$P(X=x)$= $\frac{\lambda^x}{x!} e^{-\lambda}$ x=0, 1, 2, …	λ	λ	λ>0

Means and variances for some continuous distributions

Distribution	Probability density function (p.d.f.)	Mean	Variance	Range of parameters
Uniform	$p(x)= \frac{1}{b-a} \quad a \le x \le b$	$\frac{b+a}{2}$	$\frac{1}{12}(b-a)^2$	$b>a$
Exponential	$p(x)= \lambda e^{-\lambda x}(x>0)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\lambda > 0$
Normal	$p(x)= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	μ no restriction, $\sigma^2 > 0$
Gemma	$p(x)= \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}(x>0)$	$\alpha\beta$	$\alpha\beta^2$	$\alpha > 0, \beta > 0$
Beta	$p(x)= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$ $(x>0)$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$\alpha > 0, \beta > 0$

Use MS Excel

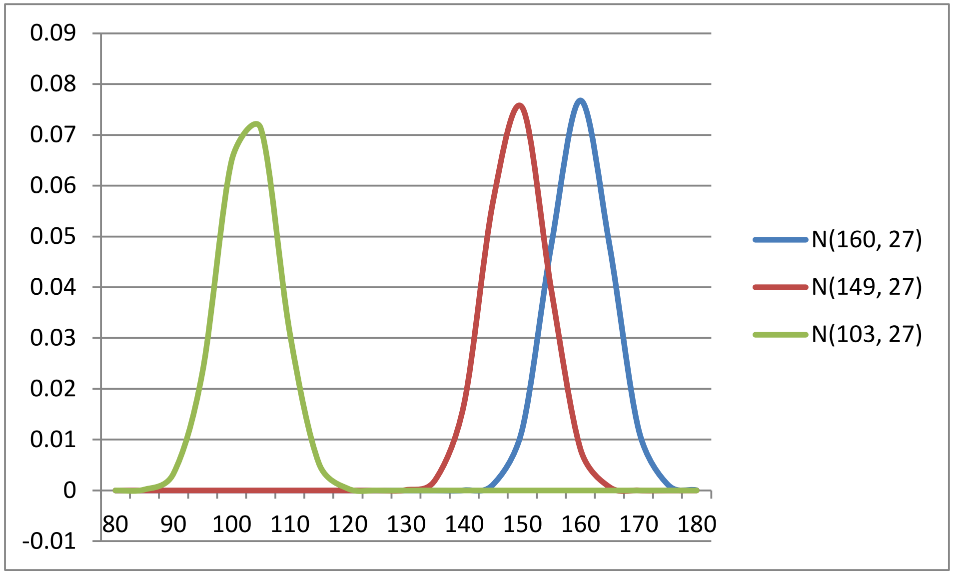
- Functions in excel for p.f. of discrete distributions and p.d.f. of continuous distributions
- Binomial distribution
  - BINOMDIST (number_s, trials, probability_s, cumulative)
- Poisson distribution
  - POISSON (x, mean, cumulative)
- Exponential distribution
  - EXPONDIST (x, lambda, cumulative)

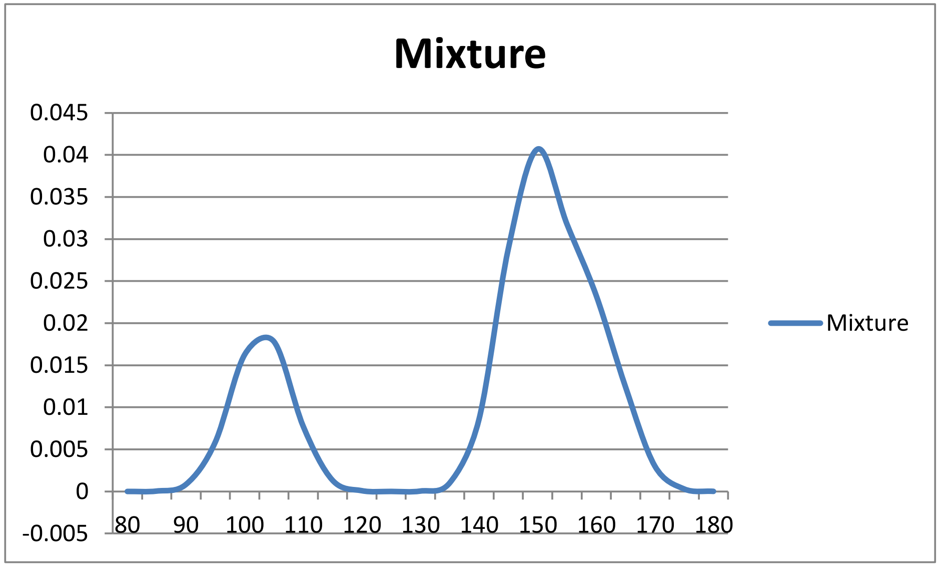
Use MS Excel

- Normal distribution
  - NORMDIST(x, mean, standard_dev, cumulative)
- Gamma distribution
  - GAMMADIST(x, alpha, beta, cumulative)
- Beta distribution
  - BETADIST(x, alpha, beta, [A], [B])

Class exercises (continued from Lecture 1)

- Draw the normal distributions of N(160, 27), N(149, 27), and N(103, 27)
- Draw the mixture distribution of N(160, 27), N(149, 27), and N(103, 27), with the proportions 0.25, 0.5 and 0.25





Class exercises: binomial and normal distributions

- Draw the bar-graph of a binomial distribution, e.g. p=0.75 and n=10
- Draw in one graph the curves of four normal distributions, e.g. N(0, 1), N(0, 4), N(3, 1), N(3, 4)

Class exercises: find the probability

$$P(X < x) = \int_{-\infty}^{x} p(t)dt$$

- X~N(0, 1), find the probability of a given interval

Interval	Probability
[-1, 1]	0.6827
[-2, 2]	0.9545
[-1.96, 1.96]	0.95
[-2.58, 2.58]	0.99
[-3, 3]	0.9973

Class exercises: find the interval

- $X \sim N(0, 1)$, find x such that $P(X<x)$=a given value

Probability	x
0.6827	0.4753
0.9545	1.6901
0.95	1.6449
0.99	2.3263
0.9973	2.7822

- $X \sim N(0, 1)$, find x such that $P(|X|<x)$=a given value

Probability	x
0.6827	1.00
0.9545	2.00
0.95	1.96
0.99	2.58
0.9973	3.00