1.0 Course introduction

Overview for MA930 Data Analysis

Day 1 Basic probability
Day 2 Basic statistics

Day 3 Frequentist statistics
Day 4 Bayesian statistics

Day 5 Basic time-series analysis
Day 6 Spectral methods for time-series analysis

Day 7 Machine-learning approaches to data analysis I
Day 8 Machine-learning approaches to data analysis II

Day 9 Class test and vivas
Day 10 Vivas continued...

Course website

https://www2.warwick.ac.uk/fac/sci/systemsbiology/staff/richardson/teaching/MA930 (https://www2.warwick.ac.uk/fac/sci/systemsbiology/staff/richardson/teaching/MA930)

Day 1 Basic probability

1.1 Probability primer
1.2 Common distributions
1.3 Distributions continued
1.4 Characteristic functions
1.5 Summary and additional questions

1.1 Probability primer

1.1.1 Rules of probability.

- Probabilities are positive and lie between 0 and 1
  $0 \leq P(A) \leq 1$
- Probability of all non-overlapping events sums to 1
- Conditionality
  $P(A|B) = P(A\&B)/P(B)$
  $P(A\&B) = P(A|B)P(B)$
- Independence
  $P(A|B) = P(A)$
- Bayes Rule
  $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$
- Total probability
  $P(A) = P(A|B)P(B) + P(A| \sim B)P(\sim B)$

1.1.1 Question: Rules of probability

Consider two events and all their possibilities

Before a test: A work or $\sim A$ play
Result of test: B pass or $\sim B$ fail

	work	**play**
pass	4/10	2/10
fail	1/10	3/10

- What are $P(\text{work})$, $P(\text{play})$, $P(\text{pass})$ and $P(\text{fail})$?
- What is $P(\text{pass} \mid \text{play})$?
- What is $P(\text{play} \mid \text{play})$?
- Are the effort put in and results independent variables?

1.1.2 Continuous and discrete probability distributions

Often events can be labelled by a

- discrete variable k with a countable number of states
- continuous variable with x

Discrete states have a probability $P(k)$

- With summation rule $1 = \sum_{\{k\}} P(k)$
- Expectation of a function $\langle h(k) \rangle = \sum_{\{k\}} h(k)P(k)$

Continuous states have a probability density $f(x)$ where $f(x)dx$ is prob. X between $x \to x + dx$

- Integrate to unity $1 = \int dxf(x)$ and units are $1/[x]$
- Expectation of a function $\langle h(x) \rangle = \int dxh(x)f(x)$

Common expectations

- Mean $\mu = \langle x \rangle$
- Variance $\sigma^2 = \langle x^2 \rangle$-$\langle x \rangle^2$

1.1.3 Cumulative distributions

- The cumulative distribution $F(x)$ is defined as the probability that the variable is less than or equal to x so that

$$F(x) = P(X \leq x)$$

- Maps both probabilities and densities onto the range 0 and 1

$$F(x) = \int_{-\infty}^{x} f(x)dx$$

- Useful for generating random numbers from any distribution using standard uniform randoms.

$F = g(X)$ can be inverted to give $X = g^{-1}(F)$

- Return to this later

1.1.4 Question: Exponential distribution example

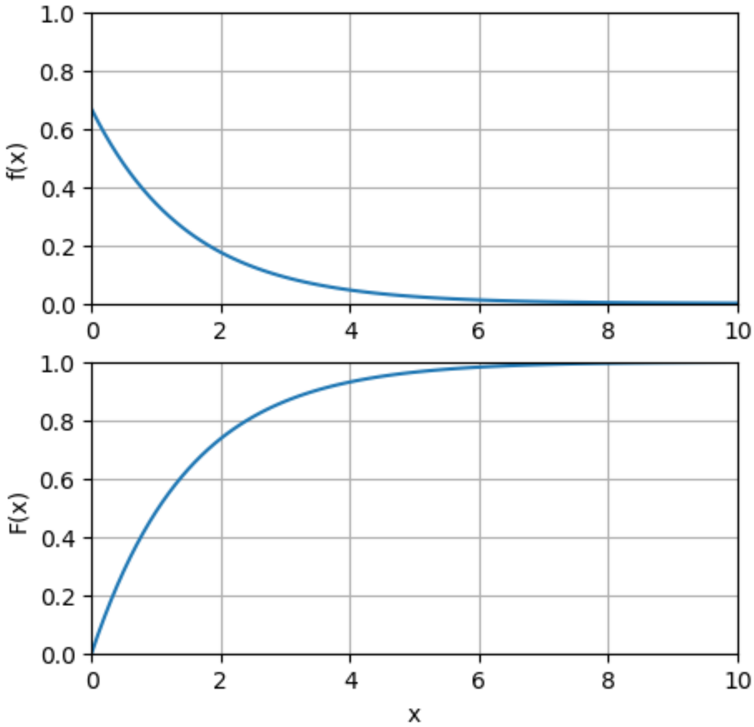
Example: Exponentially distributed random numbers obey
$f(x) = \theta(x)e^{-x/a}/a$

- What is the mean?
- What is the variance?
- What is the cumulative distribution?
- Plot the density and cumulative distributions for $a = 1.5$, one above the other, and label axes.

In [3]:

```
using PyPlot
a=1.5; x=collect(0:0.1:10)
f=exp.(-x/a)/a
F=1 .-exp.(-x/a);

figure(figsize=(5,5));
subplot(2,1,1); plot(x,f); ylabel("f(x)"); grid(); axis([0,10,0,1
])
subplot(2,1,2); plot(x,F); xlabel("x"); ylabel("F(x)"); grid(); a
xis([0,10,0,1]);
```



1.2 Useful distributions

1.2.1 Bernoulli distribution

- Discrete distribution where x is a binary random number

  $x = 1$ with probability p
  $x = 0$ with probability q

- Mean value is $\langle x \rangle = p$
- Variance. Use fact that $x^2 = x$ so that $\langle x^2 \rangle = p$.
  Hence Var$(x) = \langle x^2 \rangle - \langle x \rangle^2 = p(1-p) = pq$
- **Question:** Generate Bernoulli random numbers using the rand command and check this variance.

1.2.2 Binomial distribution

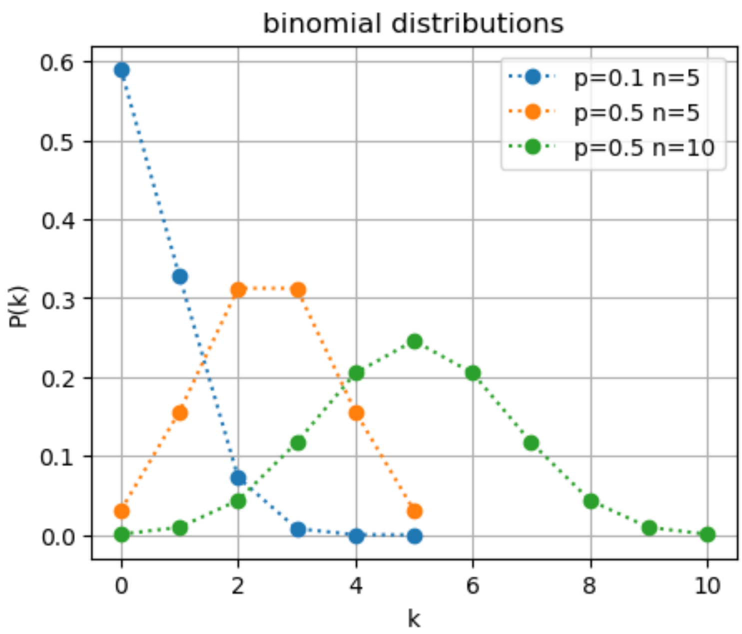
- It's a sum of n Bernouilli random variables $X = \sum_{j=1}^{n} x_j$
- Discrete distribution with $n+1$ states.
- Order unimportant so combinatorial factor is required.
- Interpret as prefactor of terms in the expansion of $(p+q)^n$ so need Pascal's Triangle

  $P(X=k) = \binom{n}{k} p^k q^{n-k}$

- Mean $\langle X \rangle = \sum_{j=1}^{n} \langle x_j \rangle = np$
- **Question:** What is the variance?

In [5]:
```
#  Binomial distribution examples
p1=0.1; q1=1-p1; n1=5;  k1=collect(0:n1); y1=binomial.(n1,k1).*(p1.^k1).*(q1.^(n1.-k1));
p2=0.5; q2=1-p2; n2=5;  k2=collect(0:n2); y2=binomial.(n2,k2).*(p2.^k2).*(q2.^(n2.-k2));
p3=0.5; q3=1-p3; n3=10; k3=collect(0:n3); y3=binomial.(n3,k3).*(p3.^k3).*(q3.^(n3.-k3));
```

In [6]:
```
figure(figsize=(5,4)); title("binomial distributions");
plot(k1,y1,":o",label="p=$p1 n=$n1"); plot(k2,y2,":o",label="p=$p2 n=$n2")
plot(k3,y3,":o",label="p=$p3 n=$n3");xlabel("k"); ylabel("P(k)");
grid(); legend();
```



1.2.3 Poisson distribution

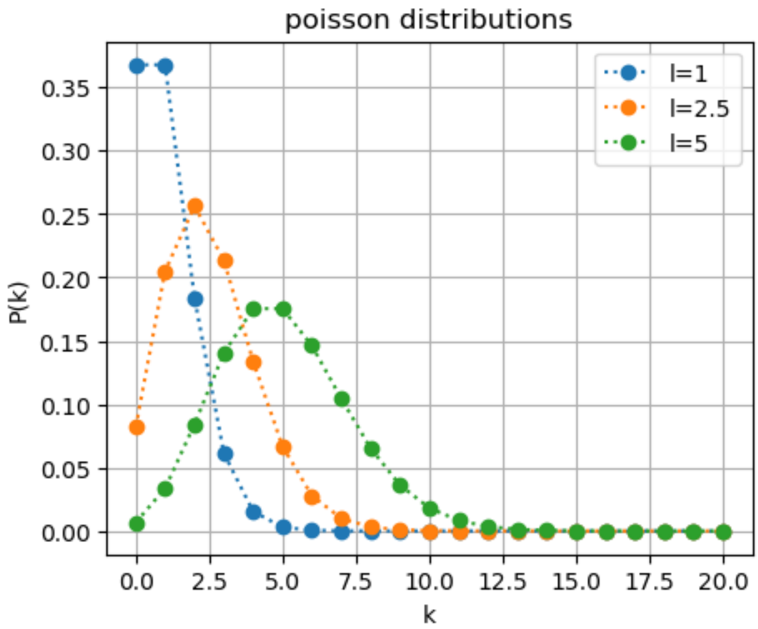
- Discrete distribution with a countable infinity of states
- Determined by parameter λ the typical number of time something happens.

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

- Mean number of events is λ
- Variance in number of events is also λ

In [9]:
```
# Poisson distribution examples
k=collect(0:20);
lam1=1;     y1=exp(-lam1)*(lam1.^k)./gamma.(k.+1)
lam2=2.5;   y2=exp(-lam2)*(lam2.^k)./gamma.(k.+1)
lam3=5;     y3=exp(-lam3)*(lam3.^k)./gamma.(k.+1);
```

In [10]:
```
figure(figsize=(5,4)); title("poisson distributions");
plot(k,y1,":o",label="l=$lam1");plot(k,y2,"o:",label="l=$lam2")
plot(k,y3,"o:",label="l=$lam3");  xlabel("k"); ylabel("P(k)"); gr
id(); legend();
```



1.2.4 Normal distribution

- Continuous distribution ubiquitous due to *Central Limit Theorem*: sums of random numbers tends to a Normal

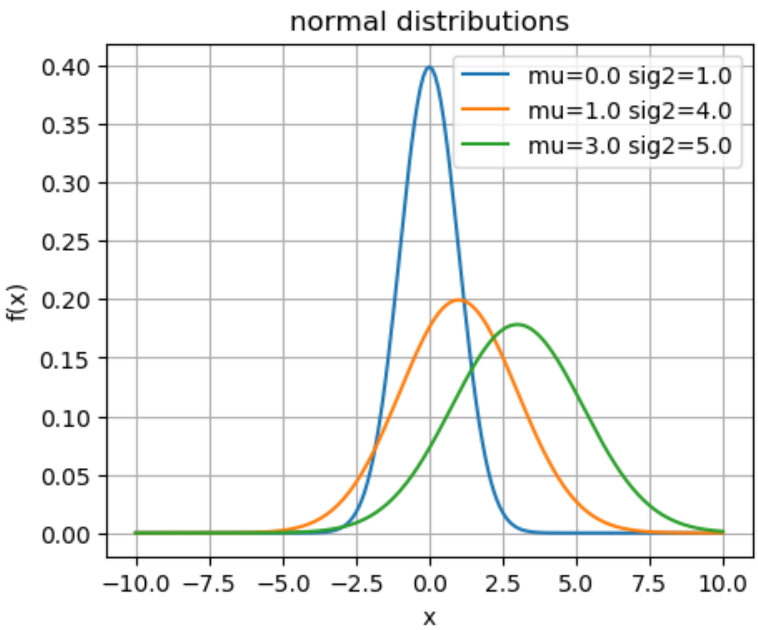
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

- Specified by the mean μ and variance σ^2.
- Standard normal: $\mu = 0$ and $\sigma^2 = 1$

- Sum of two normals with μ_1, σ_1 and μ_2 and σ_2^2 is a normal with $\mu = \mu_1 + \mu_2$ and $\sigma^2 = \sigma_1^2 + \sigma_2^2$.

In [12]:
```
# Normal distribution examples
x=collect(-10:0.1:10);
mu1=0.0; s1=1.0;       y1=exp.(-(x.-mu1).^2/(2*s1^2))/sqrt(2*pi*s1^2);
mu2=1.0; s2=2.0;       y2=exp.(-(x.-mu2).^2/(2*s2^2))/sqrt(2*pi*s2^2);
mu3=3.0; s3=sqrt(5.0); y3=exp.(-(x.-mu3).^2/(2*s3^2))/sqrt(2*pi*s3^2);
```

In [14]:
```
figure(figsize=(5,4)); title("normal distributions");
plot(x,y1,label="mu=$mu1 sig2=$(s1^2)"); plot(x,y2,label="mu=$mu2 sig2=$(s2^2)")
plot(x,y3,label="mu=$mu3 sig2=$(round(s3^2;digits=1))"); xlabel("x"); ylabel("f(x)"); grid(); legend();
```



1.2.6 Gamma distribution

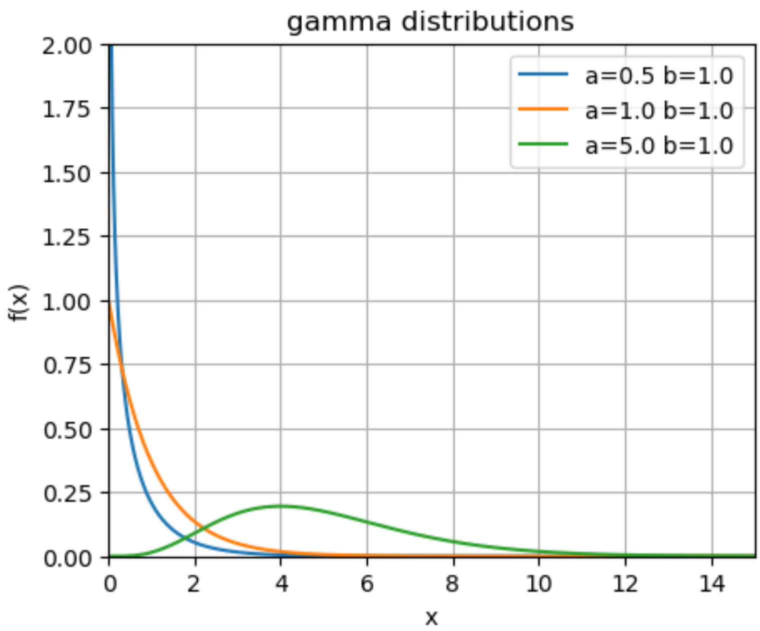
- Continuous distribution parameterised by α and β with

$$f(x) = \theta(x)\frac{1}{\Gamma(\alpha)}\beta^\alpha x^{\alpha-1}e^{-\beta x}$$

- Note the roles of α and β is setting the shape.
- **Question:** What are mean and variance in terms of α and β? HINT: use normalisation condition.

In [15]:
```
# Gamma distributions
x=collect(0.01:0.01:15.1);
a1=0.5; b1=1.0; y1=(b1^a1)*x.^(a1-1.0).*exp.(-b1*x)/gamma(a1);
a2=1.0; b2=1.0; y2=(b2^a2)*x.^(a2-1.0).*exp.(-b2*x)/gamma(a2);
a3=5.0; b3=1.0; y3=(b3^a3)*x.^(a3-1.0).*exp.(-b3*x)/gamma(a3);
```

In [16]:
```
figure(figsize=(5,4)); title("gamma distributions");
plot(x,y1,label="a=$a1 b=$b1"); plot(x,y2,label="a=$a2 b=$b2")
plot(x,y3,label="a=$a3 b=$b3"); xlabel("x"); ylabel("f(x)");
grid(); axis([0,15,0,2]); legend();
```



1.3 Distributions continued

1.3.1 Sums of random numbers
1.3.2 Multidimensional distributions
1.3.3 Marginal and conditional distributions
1.3.4 Transformations of random variables

1.3.1 Sums of random numbers

- Let the two independent random numbers x and y have distributions $f(x)$ and $g(y)$
- Then $z = x + y$ has a distribution $h(z)$ that satisfies

$h(z) = \int dx \int dy \delta(z - x - y) f(x) g(y)$

$h(z) = \int dx f(x) g(z - x)$

- This is essentially a convolution suggesting a multiplication in Fourier space.

1.3.1 Question: Gamma sum rules

- Gamma distribution $g(x) = \frac{1}{\Gamma(\alpha)} \beta^{\alpha} x^{\alpha-1} e^{-\beta x}$ for $x > 0$.
- Sum of two Gamma randoms with (α_1, β) and (α_2, β) is a Gamma random with $(\alpha_1 + \alpha_2, \beta)$
- Use the rule $g(x) = \int dx_1 \int dx_2 \delta(x - x_1 - x_2) g(x_1) g(x_2)$ to demonstrate this assertion.
- The Beta function will be of use
$B(\alpha_1, \alpha_2) = \int_0^1 du u^{\alpha_1 - 1}(1 - u)^{\alpha_2 - 1} = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}$

1.3.2 Multi-dimensional distributions

- Discrete case $P(X, Y)$ where X and Y have a countable number of states and the normalisation is

$$1 = \sum_X \sum_Y P(X, Y)$$

- Continuous case $f(x, y)$ where the normalisation is by the double integral

$$1 = \int dx \int dy f(x, y)$$

- Independence $f(x, y) = f(x)f(y)$

- Expectations of joint variables

Covariance $\sigma_{xy} = \langle (x - \mu_x)(y - \mu_y) \rangle$

Correlation $\rho_{xy} = \frac{1}{\sigma_x \sigma_y} \sigma_{xy}$ which is dimensionless

1.3.3 Marginal and conditional distributions

- Marginal distribution of X first
  $P(X) = \sum_Y P(X, Y) = \sum_Y P(X|Y)P(Y)$
- Example:

	x_1	x_2	x_3	P_y
y_1	1/16	3/16	5/16	9/16
y_2	2/16	3/16	2/16	7/16
P_x	3/16	6/16	7/16	

- For the continuous case

$$f(x) = \int dy f(x, y) = \int dy f(x|y) f(y)$$

- Note the definition of the conditional density

$$f(x|y) = f(x, y)/f(y)$$

1.3.4 Transformations of random variables

- For the continuous case. Consider a transformation from a variable Y to a new variable $X = g(Y)$.
- Can use the marginal distribution definition where the conditional density is $f_{xy}(x|y) = \delta(x - g(y))$.

$$f_x(x) = \int dy f_{x|y}(x|y) f_y(y) = f_y(y) \frac{dy}{dg}$$

- Factor dg/dy from the Dirac delta (integral over dy not dx).
- Note the transformation is $f(x)dx = f(y)dy$.
- But take care when mapping not one-to-one, like $X = Y^2$.

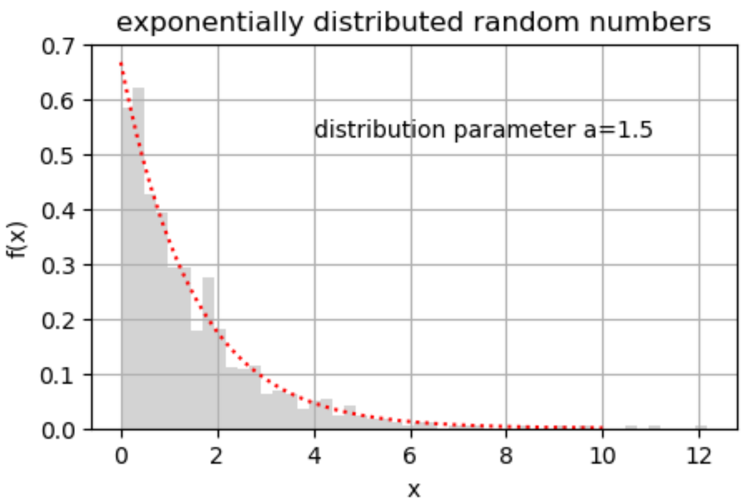
1.3.4 Question: Generating random variables

- We want to generate random variables X with a distribution $f_x(x)$.
- We can easily generate random variables Y uniformly in the range $0 \to 1$, so $f_y(y) = 1$
- What is the transformation $X = g(Y)$ required?
- What is the link to the cumulative distribution?
- Give the transformation required to generate exponentially distributed random numbers: $f(x) = \theta(x)e^{-x/a}/a$.
- Generate some exponentially distributed random numbers and compare with $f(x)$.
  NB: plt[:hist](x,50,normed=1,color="lightgray") will plot a histogram of contents of vector x.

In [19]:
```
# choose a=1.5 as before
a=1.5;
n=1000;           # number of random numbers to be generated
y=rand(n)         # generate the uniform random numbers
x=-a*log.(1 .-y) # use the transformation

xx=0:0.01:10
yy=exp.(-xx/a)/a

figure(figsize=(5,3))
plt[:hist](x,50,normed=1,color="lightgray");
plot(xx,yy,"r:"); title("exponentially distributed random numbers
")
xlabel("x"); ylabel("f(x)"); text(4.0,0.53,"distribution paramete
r a=$a");
grid("on")
```



1.4 Characteristic functions

1.4.1 Definition of the characteristic function

- Definition of the mth moment $\langle X^m \rangle$
- Moment generating function

$M(t) = \langle e^{tX} \rangle = \sum_{m=0}^{\infty} \langle X^m \rangle \frac{t^m}{m!}$

- Characteristic function takes form

$\phi_X(t) = \langle e^{itX} \rangle = \sum_{m=0}^{\infty} \langle X^m \rangle \frac{(it)^m}{m!}$

- Closely related to the moment generating function, but always exists.
- Basically a Fourier transform for continuous case.

1.4.2 Properties of the characteristic function

- Characteristic function $\phi_X(t) = \langle e^{itX} \rangle$
- Fundamental convention $\phi_X(0) = 1$
- Sum of two independent random variables $X = X_1 + X_2$

$\phi_X(t) = \langle e^{it(X_1+X_2)} \rangle = \langle e^{itX_1} \rangle \langle e^{itX_2} \rangle = \phi_{X_1}(t)\phi_{X_2}(t)$

- Obvious generalisation to sums of multiple random variables
- Useful for later. Consider a sample mean $\overline{X} = \frac{1}{n} \sum_{k=1}^{n} X_k$ from k independent samples.

$\phi_{\overline{X}}(t) = \langle e^{itX/n} \rangle^n = [\phi_X(t/n)]^n$

1.4.3 Characteristic functions of common distributions

- Example for a Bernouilli distribution with probability
  $p = 1 - q$
  $\phi_X(t) = pe^{it} + q$

- Summary for other distributions

Distribution	**characteristic function** $\phi(t)$
Bernoulli	$1 - p + pe^{it}$
Binomial	$(1 - p + pe^{it})^n$
Poisson	$e^{\lambda(e^{it}-1)}$
Normal	$e^{it\mu - \sigma^2 t^2/2}$
Gamma	$(1 - it/\beta)^{-\alpha}$

1.4.4 Calculations using characteristic functions

- Bernoulli has a characteristic function $\phi_x(t) = 1 - p + pe^{it}$

- A binomial random number is a sum of independent Bernoulli randoms $X = \sum_{j=1}^{n} x_j$

- Characteristic functions of a sum of independent randoms is a product of their individual characteristic functions.

- Hence $\phi_X(t) = \phi_x(t)^n = (1 - p + pe^{it})^n$ as expected.

1.4.4 Question: Calculations using characteristic functions

$X = X_1 + X_2$ then
$\phi_X(t) = \langle e^{it(X_1+X_2)} \rangle = \langle e^{itX_1} \rangle \langle e^{itX_2} \rangle = \phi_{X_1}(t)\phi_{X_2}(t)$

- Use this product rule for characteristic functions to derive the summation rules for:

1. Normally-distributed random numbers
2. Gamma-distributed random numbers

1.5 Summary and additional questions

Day 1 Basic probability

Additional questions

Q1.5.1 Derivation of characteristic functions
Q1.5.2 Gamma-distributed random-number generator
Q1.5.3 Using the Distributions.jl package

1.5.1 Question: Derivation of characteristic functions

- In section 1.4.3 a list of characteristic functions was given for the distributions of section 1.2. Derive these from the Bernoulli, Binomial and Poisson distributions.

1.5.2 Question: Gamma-distributed random-number generator.

- For exponentially-distributed random numbers it was possible to invert the cumulative distribution and use this to generate appropriately distributed randoms from uniformly distributed ones.
- In general it is not possible to analytically invert the cumulative distribution function.

- **Task**
- Develop a numerical method that generates gamma-distributed random numbers with shape factors $\alpha > 1$ and β.
- Check that a histogram of random numbers agrees with the original distribution.

- How well does your code work for $0 < \alpha \leq 1$?
- How might it be adapted?

1.5.3 Question: Using the Distributions.jl package

- The Distributions (https://juliastats.github.io/Distributions.jl/latest/starting.html) package provides a number of convenient functions for statistics.
- Install this in your version of Julia using Pkg.add("Distribution.jl")
- Browse the "getting started" documentation and use the package to generate some random numbers, plot the original distributions and their cumulative distributions. Here are some examples...
- Generates a named (here Normal) distribution. What are μ and σ?
  x1=-10.0:0.05:10.0;
  y1=pdf(Normal(1.0,2.0),x1);
- Generates a named cumulative distribution (here gamma). What are α and β? What is θ in this context?
  x2=0.01:0.01:5
  y2=cdf(Gamma(0.5,1.0),x2);
- Generates random numbers from a named distribution (here gamma). What are α and β?
  z=rand(Gamma(0.5, 3.0), 100)
- You can compare these with your own cumulative distribution function of the previous question.

```
In [ ]:
```