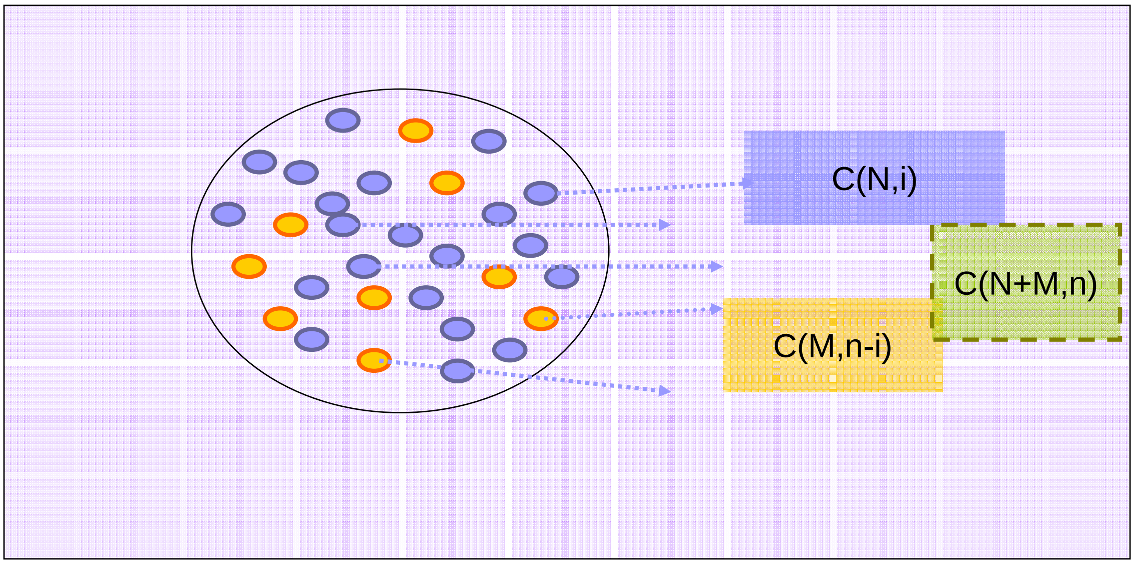
Lecture 5 (Addendum): More distributions

Objectives: to learn the following distributions and functions:

- Hypergeometric distribution
- Uniform distribution
- Multinomial distribution
- Exponential and gamma distributions
- Moment generating function (mgf)
- Chi-square, t-, and F- distributions



Hypergeometric distribution

P{X=i}=C(N , i) C(M , n-i) / C(N+M , n)

Expectation of **X**~HPG(N,M,n)

- E**X**= $\sum$ {**x** C(N , **x**) C(M , n-**x**) / C(N+M , n)}

  = $\sum$ {N C(N-1 , **x**) C(M , n-**x**)

  /[((N+M)/n) C(N-1+M , n-1)]}

  = n N/(N+M)
- Alternative approach: I_i=1 if the i-th sample is 'orange'; and I_i=0 if the i-th sample is 'blue'=>

  P{I_i=1}=N/(N+M)=E{I_i}.
- **X**= $\sum I_i$ =>E**X**= $\sum$ E{I_i}=n*N/(N+M)

Variance of **X**~HPG(N,M,n)

- Moreover, say i$\neq$j,

  P{I_i=1, I_j=1}= P{I_i=1}* P{I_j=1|I_i=1}

  =N/(N+M)* (N-1)/(N-1+M)
- Var(**X**)= $\sum$ Var(I_i)+2 $\sum_{i \neq j}$ cov(I_i, I_j)

  -- Var(I_i)=P(I_i=1)*{1- P(I_i=1)}=N/(N+M)* M/(N+M)

  because I_i is a Bernoulli rv.

  -- cov(I_i, I_j)=E(I_i I_j)- E(I_i) E(I_j) and

  E(I_i I_j)=P{I_i I_j=1}=P{I_i=1, I_j=1}

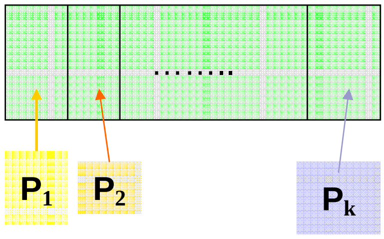
  -- Combining all these results gives

  Var(**X**)=NMn(N+M-n)/{$(N+M)^2$(N+M-1)}

The Multinomial Distribution

- Probability mass function (pmf): ($\sum p_i = 1$)

$$\frac{n!}{x_1!\cdots x_k!}p_1^{x_1}\cdots p_k^{x_k}$$



- $E(\mathbf{X_i})=nPi$,
- $Var(\mathbf{X_i})=np_i(1-p_i)$
- $Cov(\mathbf{X_i},\mathbf{X_j})= -np_ip_j$ $(i\neq j)$

Example: Virus or bacteria evolution (exercise)

- A virus consists of a sequence of DNA. It may have different sub-types. A sample of size N of this virus is supposed to have **k** different sub-types with observed sample sizes: n_1, n_2,…,n_k; with $\sum n_i = N$. The quantity C defined as

$$C= \sum n_i (n_i -1)/N(N-1),$$

is used as a measure for balanced dispersion on subtypes. Please find the variance of the statistic C. [This is problem raised by Professor CC Chang of NCHU]

The uniform distribution

- $f(\mathbf{x})=1/(\beta-\alpha)$ if $\alpha \leqq \mathbf{x} \leqq \beta$, where β and α are the two parameters. You can easily check that f(x) is integrated to 1 for x=α to x=β.
- $\mathbf{X}$~Unif(α, β); =>
  $E(\mathbf{X})=(\alpha+\beta)/2$, $Var(\mathbf{X})=(\beta-\alpha)^2/12$

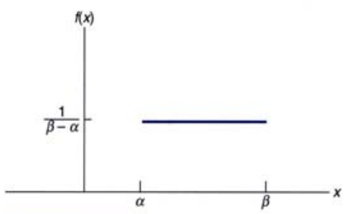


FIGURE 5.4 *Graph of f(x) for a uniform [α, β].*

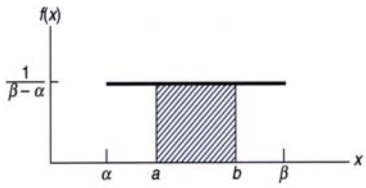
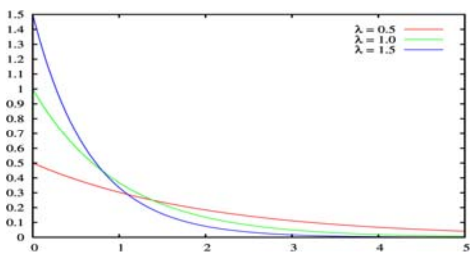
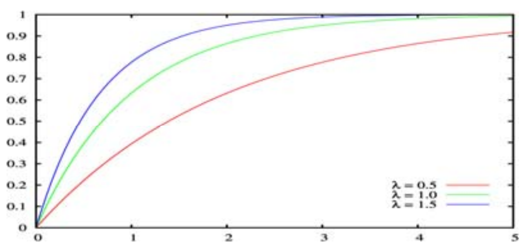


FIGURE 5.5 *Probabilities of a uniform random variable.*

Exponential distribution

- $f(\mathbf{x})=\lambda e^{-\lambda \mathbf{x}}$, if x>=0
  = 0, if x <0
- $F(\mathbf{x})= \int^{0} \lambda e^{-\lambda t}\,dt = 1- e^{-\lambda t}$, x>=0
- $E(X)=1/\lambda$, $Var(X)=1/\lambda^2$

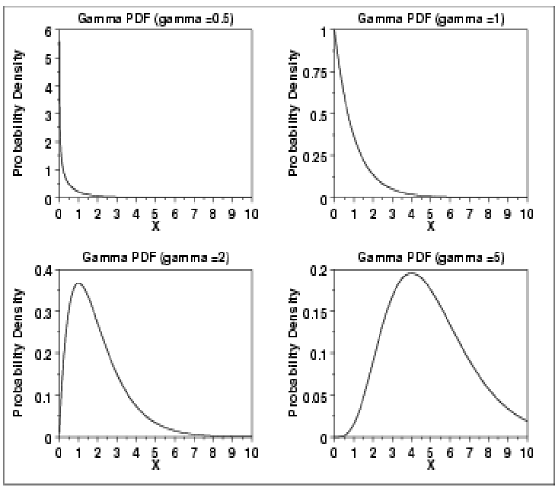


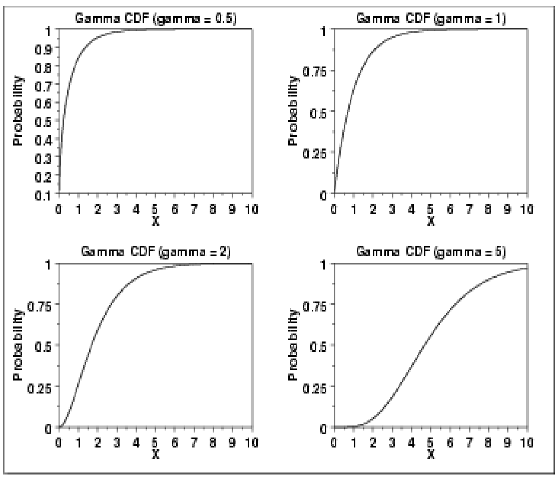


Gamma distribution

- $f(\mathbf{x}) = \lambda^{\alpha}\mathbf{x}^{\alpha-1}e^{-\lambda\mathbf{x}}/\Gamma(\alpha)$, denoted as $\Gamma(x;\alpha,\lambda)$
  where $\Gamma(\alpha)$ is a gamma function ,
  $\Gamma(\alpha) = \int^{\infty} x^{\alpha-1}e^{-x}\,dx$
- α is called the **shape** parameter
  and λ the scale parameter.
- F(x) is a direct integral of f(x), involves the incomplete gamma function. (Exercise !)
- $E(X)=\alpha/\lambda$, $Var(X)=\alpha/\lambda^2$
- **Note**: exponential distribution can be viewed as a special case of gamma distribution when $\alpha=1$.

Gamma's pdf and cdf





The moment generating function (mgf)

- $\phi(t)=E(e^{t\mathbf{X}})=\int e^{t\mathbf{x}}f(\mathbf{x})d\mathbf{x}$, which is a function of t.
- It is easy to show that:

$\phi(0)=1$; $\phi'(0)=E(\mathbf{X})$; $\phi''(0)=E(\mathbf{X}^2)$;

…. $\phi^{(k)}(0)=E(\mathbf{X}^k)$

- **Example 1**: Gaussian (Textbook, p.169)

  $\mathbf{X} \sim N(\mu, \sigma^2)$, $E(e^{t\mathbf{X}})=\exp\{\mu t+\sigma^2 t^2/2\}=\phi(t)$

  $\Rightarrow$ $\phi(0)=1$; $\phi'(0)=\mu$; $\phi''(0)=\mu^2+\sigma^2$; …

  $Var(\mathbf{X})= E(\mathbf{X}^2)- (E\mathbf{X})^2 = \sigma^2$

The moment generating function (*cont.*)

- **Example 2**: Gamma: $\mathbf{X}\sim\Gamma(\mathbf{x};\alpha,\lambda)$,

the mgf $\phi(t)=[\lambda/(\lambda-t)]^{\alpha}$;

$\phi'(t)=\alpha\ \lambda^{\alpha}/(\lambda-t)^{\alpha+1}$;

$\phi''(t)=\alpha(\alpha+1)\lambda^{\alpha}/(\lambda-t)^{\alpha+2}$;

$E(\mathbf{X})= \alpha/\lambda$; $E(x^2)= \alpha(\alpha+1)/\lambda^2$;

$Var(\mathbf{X})= E(x^2)- (Ex)^2 = \alpha/\lambda^2$

……

Homework and exercises

- Do the homework mentioned in the context of the lecture handout.
- Read sections 5.6.1 and and 5.9 to learn something about Poisson process and logistic distribution.
- Do the problems in your textbook (pages 194~200):

  Level 1: 8, 9, 16, 17, 21, 30, 34, 37, 44, 47

  Level 2: 19, 20, 29

Homework and exercises (cont.)

- Show that, if $X_1 \sim \Gamma(\alpha_1, \lambda)$ and $X_2 \sim \Gamma(\alpha_2, \lambda)$; then the random variable $Y = X_1 + X_2 \sim \Gamma(\alpha_1 + \alpha_2, \lambda)$.
- For a gamma function, it is easy to show:

  (1) $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$; and

  (2) $\Gamma(n) = (n-1)!$ for an integer value "n".

  (3) Moreover, $\Gamma(1/2) = \sqrt{\pi}$