Computational Physics

Examples of Probability Functions

Korea University
Eunil Won

Binomial distribution

• Binomial distribution

Consider a series of *N* independent trials with only two possible outcomes. When the desired outcome of one trial is *p*, the probability of finding *n* desired outcome from *N* trials is governed by the binomial distribution.

The probability to have *n* desired outcomes in *N* events is

$$f(n; N, p) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$

The expectation and the variance are

Can you prove them?

$$E[n] = \sum_{n=0}^{N} n \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} = Np$$

$$V[n] = E[n^2] - (E[n])^2 = Np(1-p)$$

Binomial distribution - proof 1

The expectation value can be found as

$$
\begin{aligned}
E[n] &= \sum_{n=0}^{N} n \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} \\
&= Np \sum_{n=1}^{N} \frac{(N-1)!}{(n-1)!\big((N-1)-(n-1)\big)!} p^{(n-1)} (1-p)^{(N-1)-(n-1)} \\
& \quad \text{with } m = n-1, \quad M = N-1, \\
&= Np \sum_{m=0}^{M} \frac{M!}{m!(M-m)!} p^m (1-p)^{M-m} \\
&= Np
\end{aligned}
$$

Binomial distribution - proof 2

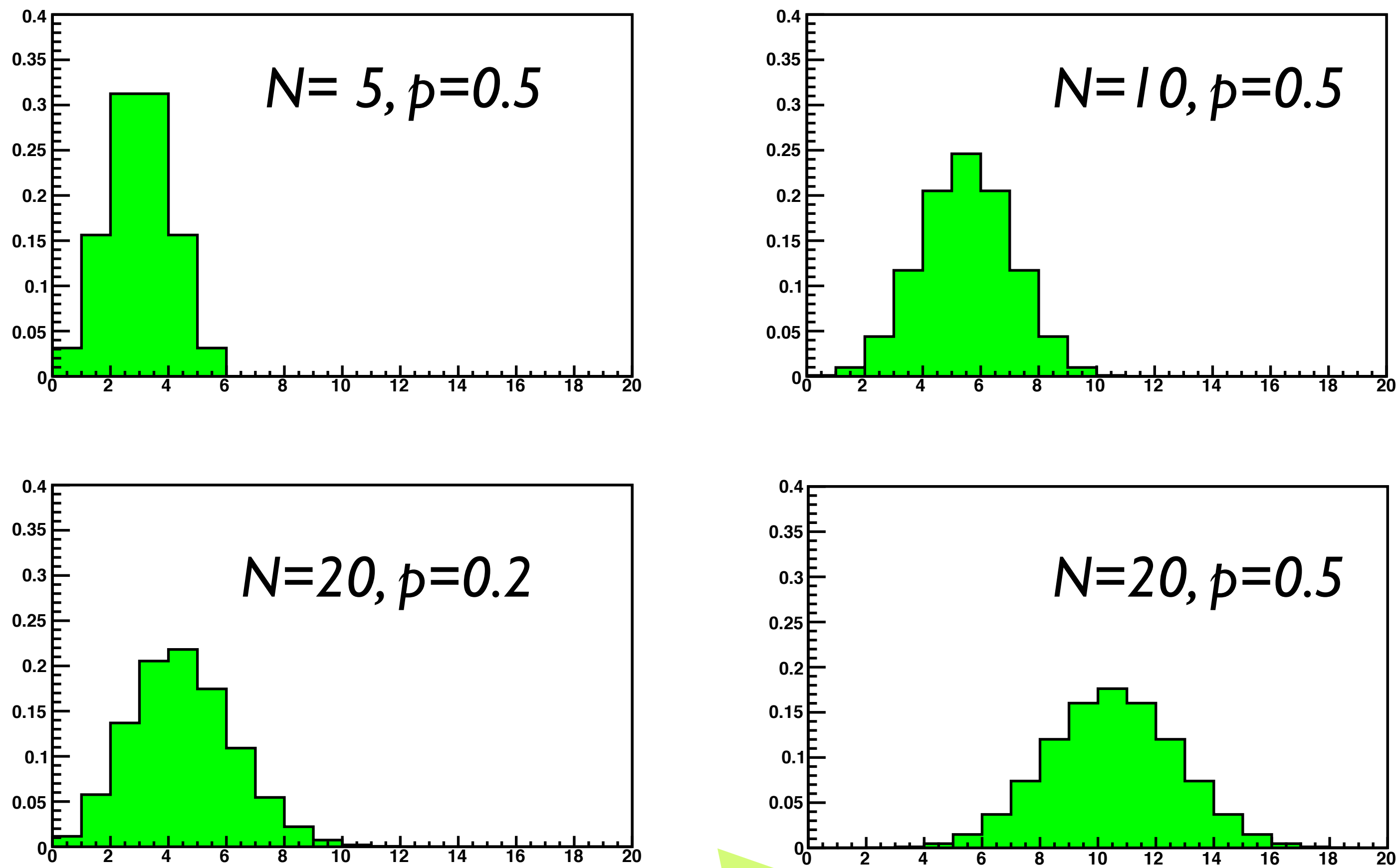
The variance can be found as

$$E[n^2] = E[n(n-1)+n] = E[n(n-1)] + E[n]$$

$$\begin{aligned} E[n(n-1)] &= \sum_{n=0}^{N} n(n-1)\frac{N!}{n!(N-n)!}p^n(1-p)^{N-n} \\ &= \sum_{n=2}^{N} \frac{N!}{(n-2)!(N-n)!}p^n(1-p)^{N-n} \\ &= N(N-1)p^2 \sum_{n=2}^{N} \frac{(N-2)!}{(n-2)!\big((N-2)-(n-2)\big)!}p^{n-2}(1-p)^{(N-2)-(n-2)} \\ &= N(N-1)p^2 \end{aligned}$$

$$\therefore V[n] = E[n^2] - (E[n])^2 = N(N-1)p^2 + Np - (Np)^2 = Np(1-p)$$

Binomial distribution - example

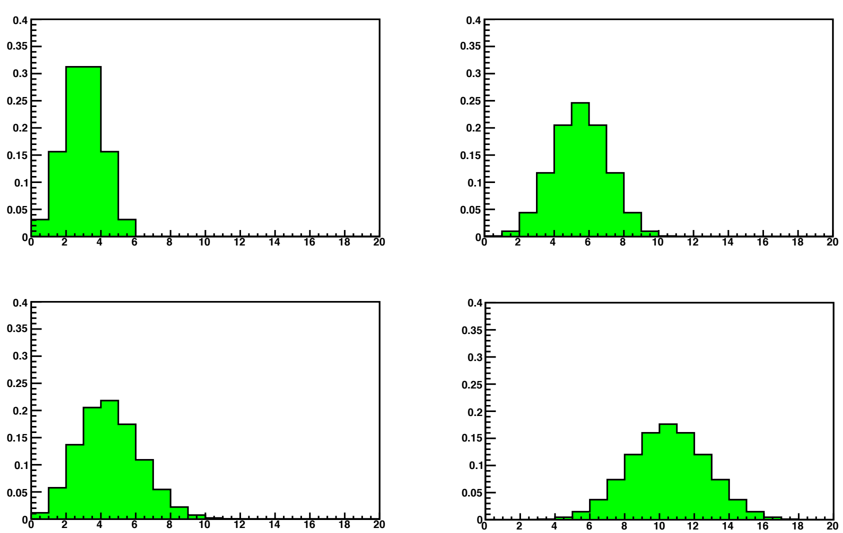


Can you write ROOT C++ code that draws this?

Binomial distribution

- the source code

The function f1 returns
the binomial distribution



```
#include <iostream>
//
// Drawing a binomial function
//
// E. Won (eunil@hep.korea.ac.kr)

Double_t f1(Int_t n, Int_t N, Double_t p)
{
  Double_t fac = TMath::Power(p,n)*TMath::Power((1.0-p),N-n);
  for (Int_t i=1;i<=n;i++,N--)
  {
    fac /= i;
    fac *= N;
  }
  return fac;
}

void binomial()
{
  gROOT->SetStyle("Plain");
  gROOT->ForceStyle();

  TCanvas* c = new TCanvas("c","",200,10,600,400);
  c->Divide(2,2);
  h1 = new TH1F("h1","",20,0.0,20); h1->SetMaximum(0.4);
  h2 = new TH1F("h2","",20,0.0,20); h2->SetMaximum(0.4);
  h3 = new TH1F("h3","",20,0.0,20); h3->SetMaximum(0.4);
  h4 = new TH1F("h4","",20,0.0,20); h4->SetMaximum(0.4);
  h1->SetFillColor(3);h2->SetFillColor(3);
  h3->SetFillColor(3);h4->SetFillColor(3);
  for (Int_t i=0;i<20;i++)
  {
    h1->Fill(i,f1(i, 5,0.5)); h2->Fill(i,f1(i,10,0.5));
    h3->Fill(i,f1(i,20,0.2)); h4->Fill(i,f1(i,20,0.5));
  }
  c->cd(1); h1->Draw(); c->cd(2); h2->Draw();
  c->cd(3); h3->Draw(); c->cd(4); h4->Draw();
}
```

Binomial distribution - example

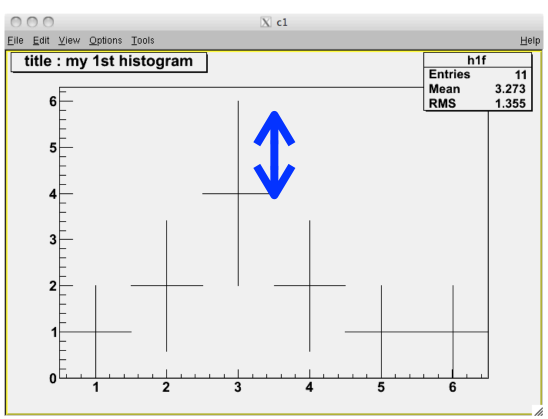
- Histogram

Suppose we are interested in one particular bin of a compound histogram. Then "success" may correspond to getting an entry in this particular bin, and "failure" corresponds to an entry in any other bin of the histogram - so it is related with the binomial distribution.

The probability p for a success is $p = n/N$ when n and N are entries in the bin and the total number of events in the histogram.

$$V[n] = Np(1-p) = N\frac{n}{N}\Big(1-\frac{n}{N}\Big) = n\Big(1-\frac{n}{N}\Big)$$

$$\sigma[n] = \sqrt{n\Big(1-\frac{n}{N}\Big)} \approx \sqrt{n} \quad \text{for } p \to 0 \quad (\text{or} \quad N \to \infty)$$



We did this when we discussed histogram errors!

σ_{y_i} : error of y_i

$$\sigma_{y_i} = \sqrt{y_i}$$

Half size of the vertical bar shows the "error of y_i"

Multinomial distribution

- Multinomial distribution

Generalization of the binomial distribution with m different possible outcomes. For a particular outcome has the probability p_i, and of course $\sum_{i=1}^{m} p_i = 1$

$$f(n_1, ..., n_m; N, p_1, ..., p_m) = \frac{N!}{n_1! n_2! .. n_m!} p_1^{n_1} p_2^{n_2} ... p_m^{n_m}$$

The expectation and the variance are

$$E[n_i] = Np_i \quad \text{and} \quad V[n_i] = Np_i(1 - p_i)$$

The covariance becomes

$$\text{cov}[n_i, n_j] = -Np_i p_j$$

ex) For three possible outcomes

$$f(n_i, n_j; N, p_i, p_j) = \frac{N!}{n_i! n_j! (N - n_i - n_j)!} p_i^{n_i} p_j^{n_j} (1 - p_i - p_j)^{N - n_i - n_j}$$

Poisson Distribution

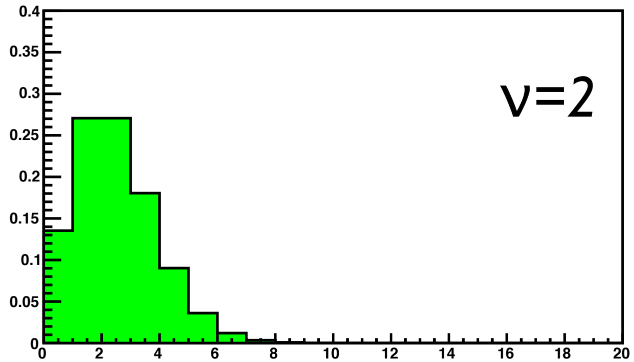
- Poisson distribution

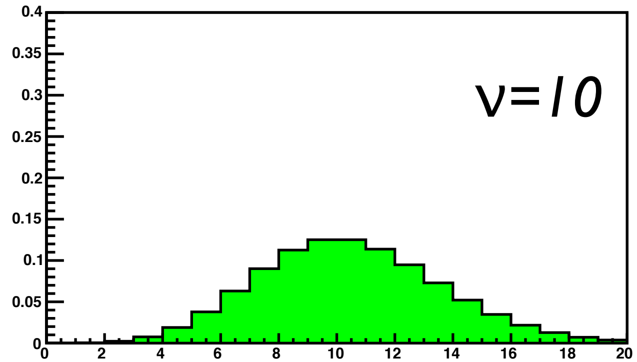
Limiting case of the binomial distribution when N becomes large, p very small but the product Np (i.e. the expectation value) remains a finite value ν

$$f(n;\nu) = \frac{\nu^n}{n!} e^{-\nu}$$

The expectation and the variance are

$$E[n] = \sum_{n=0}^{\infty} n \frac{\nu^n}{n!} e^{-\nu} = \nu \qquad V[n] = \sum_{n=0}^{\infty} (n-\nu)^2 \frac{\nu^n}{n!} e^{-\nu} = \nu$$





Poisson Distribution

- Poisson distribution

Limiting case of the binomial distribution when *N* becomes large, *p* very small but the product *Np* (i.e. the expectation value) remains a finite value ν - proof

$Np = \nu$ and Stirring's formula gives $N! \approx \sqrt{2\pi N} N^N e^{-N}$

so,

$$
\begin{aligned}
\frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} &\approx \frac{1}{n!} \frac{\sqrt{2\pi N} N^N e^{-N}}{\sqrt{2\pi(N-n)}(N-n)^{(N-n)} e^{-(N-n)}} \left(\frac{\nu}{N}\right)^n \left(1-\frac{\nu}{N}\right)^{(N-n)} \\
&\approx \frac{1}{n!} \frac{N^{(N-n)}}{(N-n)^{(N-n)} e^{-(N-n)}} \nu^n \left(1-\frac{\nu}{N}\right)^{(N-n)} \\
&\approx \frac{1}{n!} \frac{1}{\left(1-\frac{n}{N}\right)^{(N-n)}} \frac{1}{e^n} \nu^n \left(1-\frac{\nu}{N}\right)^{(N-n)} \\
&\approx \frac{e^{-n} \nu^n}{n!}
\end{aligned}
$$

Uniform Distribution

• Uniform distribution

The uniform p.d.f. for the continuous variable is defined by

$$f(x;\alpha,\beta) = \begin{cases} \frac{1}{\beta-\alpha} & \alpha \le x \le \beta \\ 0 & \text{otherwise} \end{cases}$$

The expectation and the variance are

$$E[x] = \int_\alpha^\beta \frac{x}{\beta-\alpha} dx = \frac{1}{2}(\alpha+\beta)$$

$$V[x] = \int_\alpha^\beta [x - \frac{1}{2}(\alpha+\beta)]^2 \frac{1}{\beta-\alpha} dx = \frac{1}{12}(\beta-\alpha)^2$$

Exponential Distribution

- Exponential distribution

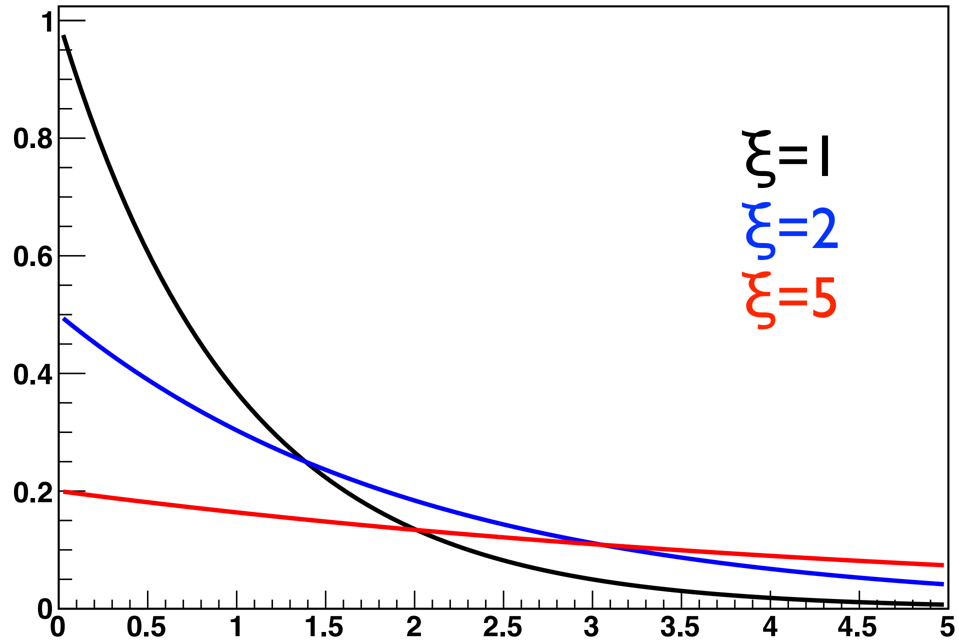
The exponential p.d.f. for the continuous variable is defined by

$$f(x,\xi) = \frac{1}{\xi} e^{-x/\xi} \qquad (0 \le x < \infty)$$

The expectation and the variance are

$$E[x] = \frac{1}{\xi} \int_0^\infty x e^{-x/\xi} dx = \xi$$

$$V[x] = \frac{1}{\xi} \int_0^\infty (x-\xi)^2 e^{-x/\xi} dx = \xi^2$$



Gaussian Distribution

- Gaussian distribution

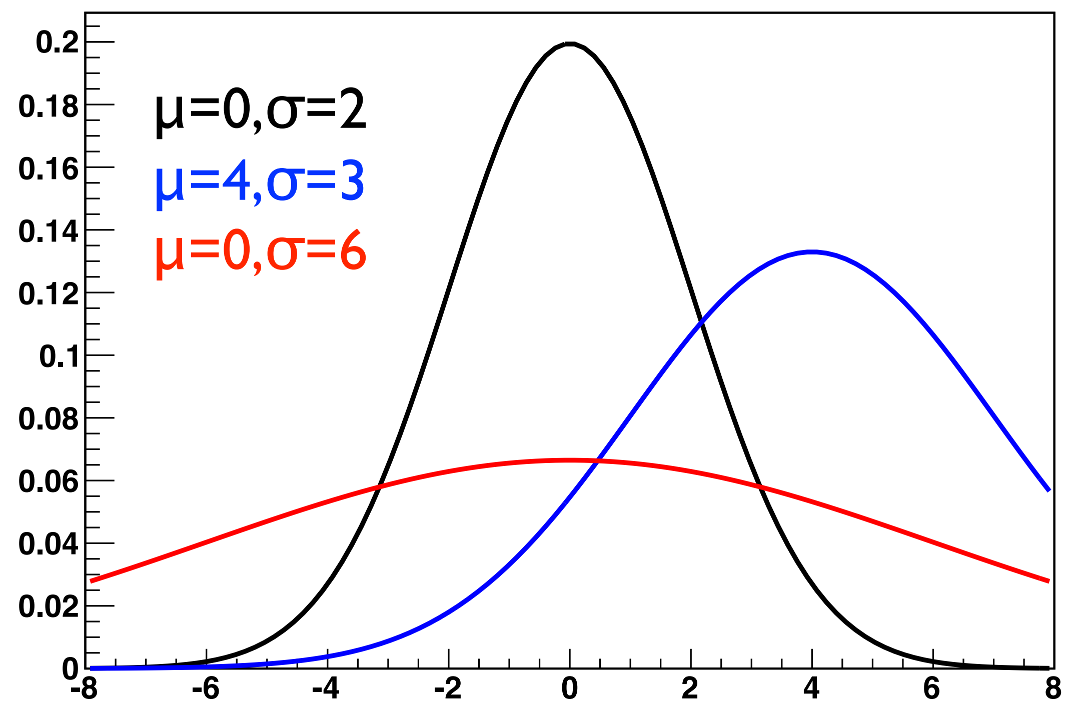
The gaussian p.d.f. for the continuous variable is defined by

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

The expectation and the variance are

$$\begin{aligned} E[x] &= \mu \\ V[x] &= \sigma^2 \end{aligned}$$

Can you prove them?



N-dim. Gaussian Distribution

- Generalized N-dimensional gaussian distribution

The N-dim. gaussian p.d.f. for the continuous variables **x** is defined by

$$f(\mathbf{x};\boldsymbol{\mu},V) = \frac{1}{(2\pi)^{N/2}|V|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T V^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

where **x** and **μ** are column vectors containing $x_1,...,x_n$ and $\mu_1,...,\mu_n$, $\mathbf{x}^T$ and $\boldsymbol{\mu}^T$ are the corresponding row vectors, and $|V|$ is the determinant of a symmetric *NxN* matrix *V*.

The expectation and the (co)variance are

$$E[x_i] = \mu_i, \quad V[x_i] = V_{ii}, \quad \mathrm{cov}[x_i, x_j] = V_{ij}$$

The two dimensional example becomes

$$f(x_1,x_2;\mu_1,\mu_2,\sigma_1,\sigma_2,\rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

$$\times \quad \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]\right\}$$

where $\rho = \mathrm{cov}[x_1,x_2]/(\sigma_1\sigma_2)$ is the correlation coefficient

Chi-square Distribution

- The χ^2(chi-square) distribution

The χ^2(chi-square) distribution of the continuous variable z is defined by

$$f(z;n) = \frac{1}{2^{n/2}\Gamma(n/2)} z^{n/2-1} e^{-z/2}, \quad n = 1, 2, \ldots \quad (0 \leq z < \infty)$$

where the parameter n is called the **number of degrees of freedom** and the gamma function is defined by

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

The expectation and the variance are found to be

$$\begin{aligned} E[z] &= \int_0^\infty z \frac{1}{2^{n/2}\Gamma(n/2)} z^{n/2-1} e^{-z/2} \, dz = n, \\ V[z] &= \int_0^\infty (z-n)^2 \frac{1}{2^{n/2}\Gamma(n/2)} z^{n/2-1} e^{-z/2} \, dz = 2n \end{aligned}$$

Chi-square Distribution

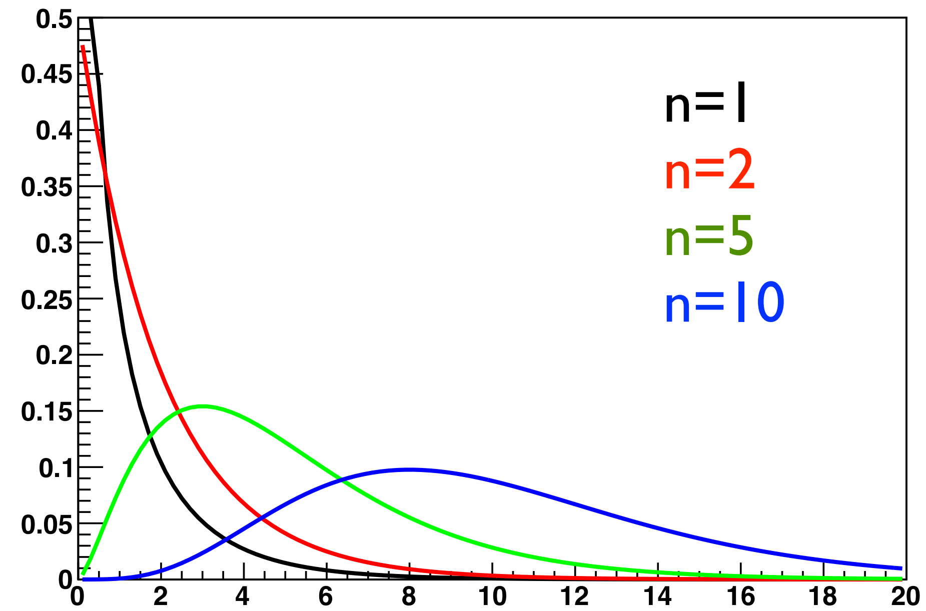
```
//
// E. Won (eunil@hep.korea.ac.kr)
//
// gammln from Numerical Recipes

float gammln(float xx)
{
        double x,y,tmp,ser;
        static double cof[6]={76.18009172947146,-86.50532032941677,
                24.01409824083091,-1.231739572450155,
                0.1208650973866179e-2,-0.5395239384953e-5};
        int j;

        y=x=xx;
        tmp=x+5.5;
        tmp -= (x+0.5)*log(tmp);
        ser=1.000000000190015;
        for (j=0;j<=5;j++) ser += cof[j]/++y;
        return -tmp+log(2.5066282746310005*ser/x);
}

Double_t f1(Double_t *x, Double_t *n)
{
  Double_t lnf = TMath::Log(1) - n[0]*0.5*TMath::Log(2.0) - gammln(n[0]*0.5)
              + (n[0]*0.5-1.0)*TMath::Log(x[0]) - x[0]*0.5;
  return TMath::Exp(lnf);
}

void chisquare()
{
  gROOT->SetStyle("Plain");
  gROOT->ForceStyle();
  TH2F *h2  = new TH2F("h2","",2,0.0,20.0,2,0.0,0.5); h2->Draw();
  TF1 *fun1 = new TF1("fun1",f1,0.0,20.0,1);
  TF1 *fun2 = new TF1("fun2",f1,0.0,20.0,1);
  TF1 *fun3 = new TF1("fun3",f1,0.0,20.0,1);
  TF1 *fun4 = new TF1("fun4",f1,0.0,20.0,1);
  fun1->SetParameter(0, 1); fun1->Draw("LSAME");
  // rest of lines skipped ...
```



Can you understand how it is done? (I took the logarithm first!)