Non-uniform distributions

Non-uniform distributions

Most situation in physics – random numbers with non-uniform distribution

- radioactive decay
- experiments with different types of distributions
- …

Principal idea: Generating non-uniform random number distributions with a uniform random number generators

Probability Distribution Functions PDF

	Discrete PDF	continuous PDF
Domain	$\{x_1, x_2, x_3, \dots, x_N\}$	$[a, b]$
probability	$p(x_i)$	$p(x)dx$
Cumulative	$P_i = \sum_{l=1}^{i} p(x_l)$	$P(x) = \int_a^x p(t)dt$
Positivity	$0 \leq p(x_i) \leq 1$	$p(x) \geq 0$
Positivity	$0 \leq P_i \leq 1$	$0 \leq P(x) \leq 1$
Monotonuous	$P_i \geq P_j$ if $x_i \geq x_j$	$P(x_i) \geq P(x_j)$ if $x_i \geq x_j$
Normalization	$P_N = 1$	$P(b) = 1$

As an example, consider the tossing of two dice, which yields the following possible values

$$[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12].$$

These values are called the *domain*. To this domain we have the corresponding *probabilities*

$$[1/36, 2/36/3/36, 4/36, 5/36, 6/36, 5/36, 4/36, 3/36, 2/36, 1/36].$$

Expectation values

1. Discrete PDF

$$E[x^k] = \langle x^k \rangle = \frac{1}{N}\sum_{i=1}^{N} x_i^k p(x_i)$$

2. Continuous PDF

$$E[x^k] = \langle x^k \rangle = \int_a^b x^k p(x) dx$$

3. Expectation of function $f(x)$

$$E[f^k] = \langle f^k \rangle = \int_a^b dx\, f^k(x) p(x)$$

4. Variance

$$\sigma_f^2 = E[f^2] - (E[f])^2 = \langle f^2 \rangle - \langle f \rangle^2$$

Uniform distribution

The uniform PDF

$$p(x) = \frac{1}{b-a}\theta(b-x)\theta(x-a)$$

where $\theta(x) = 1$ if $x \geq 0$ and $\theta(x) = 0$ if $x < 0$.

For $a = 0,\ b = 1$ this distribution gives $p(x) = 1$ for $1 \geq x \geq 0$ and 0 everywhere else. It forms the basis for all generators of random numbers.

Exponential distribution

The exponential PDF

$$p(x) = \alpha e^{-\alpha x}$$

gives positive probabilities for $x > 0$ with mean value

$$\mu = \int_0^\infty dx\, xp(x) = \int_0^\infty dx\, x\alpha e^{-\alpha x} = \frac{1}{\alpha}$$

and variance

$$\sigma = \int_0^\infty dx\, x^2 p(x) - \mu^2 = \int_0^\infty dx\, x\alpha e^{-\alpha x} = \frac{2}{\alpha^2} - \frac{1}{\alpha^2} = \frac{1}{\alpha^2}$$

Normal distribution

The normal PDF

$$p(x) = \frac{1}{b\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2b^2}}$$

gives positive probabilities for any x with mean value

$$\mu = \int_{-\infty}^{\infty} dx\, x p(x) = \frac{1}{b\sqrt{2\pi}} \int dx\, x e^{-\frac{(x-a)^2}{2b^2}} = \frac{1}{b\sqrt{2\pi}} \int dx\, (x+a) e^{-\frac{x^2}{2b^2}} = a$$

Gaussian integral

$$\int_{-\infty}^{\infty} dx\, e^{-x^2} = \sqrt{\pi} \Rightarrow \int dx\, e^{-\frac{x^2}{2b^2}} = b\sqrt{2\pi}$$

Differentiate with respect to b $\Rightarrow$

$$\int dx\, x^2 e^{-\frac{x^2}{2b^2}} = b^3\sqrt{2\pi}$$

Normal distribution

$\Rightarrow$ the variance is

$$\sigma^2 = \int_{-\infty}^{\infty} dx\,(x-\mu)^2 p(x) = \frac{1}{b\sqrt{2\pi}} \int dx\,(x-a)^2 e^{-\frac{(x-a)^2}{2b^2}} = \frac{1}{b\sqrt{\pi}} \int dx\,x^2 e^{-\frac{x^2}{2b^2}} = b^2$$

It is convenient to introduce standard normal distribution with $\mu = a = 0$ and variance $\sigma^2 = 1$

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2}}$$

The uniform and exponential distributions have simple cumulative functions but the normal distribution have cumulative special function called error function erf(x)

$$\text{erf}(x) = \int_{-\infty}^{x} dt\, e^{-t^2}$$

Discrete random variables

Suppose we measure some discrete random variable X. If the result of the measurement x_1 comes with probability p_1, the result x_2 with probability p_2 e

$$\mu \equiv \langle x \rangle = \sum_{i=1}^{n} p_i x_i$$

where p_i is the probability mass function.
All p_i must be positive and normalized

$$\sum_{x=0}^{n} p_i = 1 \qquad \text{normalization}$$

Example: for dice game $p_1 = p_2 = \ldots = p_6 = \frac{1}{6}$
Variance

$$\sigma^2 \equiv \langle (x-\mu)^2 \rangle = \sum_{i=1}^{n} p_i (x_i - \mu)^2$$

Binomial distribution

Suppose we are making n independent success/failure experiments with probability of success being s. The probability of exactly x successes in n trials is given by the binomial distribution

$$p(x) = \frac{n!}{x!(n-x)!} s^x (1-s)^{n-x} \qquad x = 1, 2 \ldots n$$

Normalization

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

Newton's binomial

$$(a+b)^n = \sum_{m=1}^{n} \frac{n!}{m!(n-m)!} a^m b^{n-m}$$

Mean $\mu = ns$ and variance $\sigma^2 = ns(1-s)$

Binomial mean and variance

Mean

$$\mu = \sum_{x=0}^{n} xp(x) = \sum_{x=1}^{n} \frac{n!s^x(1-s)^{n-x}}{(x-1)!(n-x)!} = ns\sum_{x=1}^{n} \frac{(n-1)!s^{x-1}(1-s)^{(n-1)-(x-1)}}{(x-1)![(n-1)-(x-1)]!}$$

$$= ns\sum_{y=0}^{n-1} \frac{(n-1)!s^y(1-s)^{(n-1)-y}}{y![(n-1)-y]!} = ns[s+(1-s)]^{n-1} = ns$$

Variance

$$\sigma^2 = \sum_{x=0}^{n}(x^2-\mu^2)p(x) = \sum_{x=1}^{n} \frac{(x-1+1)n!s^x(1-s)^{n-x}}{(x-1)!(n-x)!} - n^2s^2$$

$$= n(n-1)s^2\sum_{x=2}^{n} \frac{(n-2)!s^{x-2}(1-s)^{(n-2)-(x-2)}}{(x-2)![(n-2)-(x-2)]!} + \sum_{x=1}^{n} \frac{n!s^x(1-s)^{n-x}}{(x-1)!(n-x)!} - n^2s^2$$

$$= n(n-1)s^2\sum_{y=0}^{n-2} \frac{(n-2)!s^y(1-s)^{(n-2)-y}}{y![(n-2)-y]!} + ns - n^2s^2 = ns(1-s)$$

Poisson distribution

Consider binomial distribution at $n \to \infty$, $s = \frac{\lambda}{n} \to 0$ (with λ fixed) and finite x.

$$p(\lambda) = \lim_{n\to\infty} \frac{n!}{x!(n-x)!} \frac{\lambda^x}{n^x} \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{\lambda^x}{x!} \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n \lim_{n\to\infty} \frac{n(n-1)\ldots(n-x+1)}{n^x} \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-x} = \frac{\lambda^x}{x!} e^{-\lambda}$$

$\Rightarrow$ Poisson distribution $p(x) = \frac{\lambda^x}{x!} e^{-\lambda}$

Normalization

$$\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} e^{-\lambda} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$$

Mean

$$\mu = \sum_{x=1}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

Variance

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

Multivariable distributions

Assume that we have two sets of measurements X_1 and X_2 with probability p_{ij} of getting x_{1i} in the first measurement and x_{2j} in the second.

Normalization: $\sum_{i,j=1}^{N} p_{ij} = 1$ (N is called sample size).

The means $\mu_1 = \langle x_1 \rangle$ and $\mu_2 = \langle x_2 \rangle$ are

$$\mu_1 = \langle x_1 \rangle = \sum_{i,j=1}^{N} x_{1i} p_{ij}, \quad \mu_2 = \langle x_2 \rangle = \sum_{i,j=1}^{N} p_{ij} x_{2j}$$

A measure of strength of correlation between X_1 and X_2 is given by the covariance

$$\mathrm{cov}(X_1, X_2) \equiv \langle (x_1 - \mu_1)(x_2 - \mu_2) \rangle = \sum_{i,j=1}^{N} p_{ij}(x_{1i} - \mu_1)(x_{2j} - \mu_2)$$

$$= \sum_{i,j=1}^{N} p_{ij} x_{1i} x_{2j} - \mu_1 \mu_2 = \langle x_1 x_2 \rangle - \mu_1 \mu_2$$

Central Limit Theorem

Suppose we have a sequence of independent measurements (random variables) with the same PDF $p(x)$ with mean μ and variance σ^2. Suppose we are interested in the sample average of these measurements

$$z_N \equiv \frac{x_1 + x_2 + \ldots + x_N}{N}$$

where each x_i is the result of $i - th$ measurement.

Q:What is then the PDF of a new variable z?
A: It is obvious that the mean μ of z is same as mean $\mu = \langle x_i \rangle$ of each x_i.
Less obvious:

$$\sigma_N^2 = \frac{\sigma^2}{N} \qquad \text{(Central Limit Theorem)}$$

The central limit theorem states that as N gets larger, the distribution of the difference between the sample average z_N and its limit μ, when multiplied by the factor $\sqrt{N}$ approximates the normal distribution with mean 0 and variance σ^2.

Proof of central limit theorem

The probability to get x_1 on the first measurement is $p(x_1)$, x_2 on the second measurement is $p(x_2)$ etc.

Since measurements are independent, the probability to get x_1 on the first measurement and x_2 on the second measurement .... and x_N on the N-th measurement is $p(x_1)p(x_2)....p(x_N)$.

Since we are interested only in the sum $x_1 + x_2 + ...x_N$ and not how it was obtained from individual measurements we need to sum over all possibilities with the constraint that $x_1 + x_2 + ... + x_N = z$.
Mathematically it can be expressed as

$$\tilde{p}(z) = \sum p(x_1)p(x_2)....p(x_N)\delta\left(z = \frac{x_1 + x_2 + ...x_N}{N}\right)$$

where $\delta\left(z = \frac{x_1+x_2+....x_N}{N}\right) = 1$ if $z = \frac{x_1+x_2+....x_N}{N}$ and 0 otherwise.

For continuous random variables

$$\tilde{p}(z) = \int dx_1 p(x_1) \int dx_2 p(x_2) .... \int dx_N p(x_N)\ \delta\left(z - \frac{x_1 + x_2 + ...x_N}{N}\right)$$

where Dirac δ-function enforces the constraint that the mean is z.

The definition of Dirac δ-function $\delta(x)$ is ∞ at $x = 0$ and 0 everywhere else such that

$$\int dx\, f(x)\delta(x) = f(0) \quad \text{for any } f(x).$$

The δ-function can be expressed as

$$\delta(x) = \int \frac{dq}{2\pi}\ e^{ipx}$$

and then the expression for $\tilde{p}(z)$ factorizes into a product of one-dimensional integrals

$$\Rightarrow \tilde{p}(z) = \int \frac{dq}{2\pi}\ e^{iqz} \int dx_1 p(x_1) \int dx_2 p(x_2) .... \int dx_N p(x_N)\ e^{i\frac{q}{N}(x_1 + x_2 + ... + x_n)}$$

$$= \int \frac{dq}{2\pi} e^{iq(z-\mu)} \left[\int dx p(x)\ e^{\frac{iq(x-\mu)}{N}}\right]^N$$

Now we consider the individual integral

$$\int dx\, p(x)\, e^{\frac{iq(x-\mu)}{N}}$$

Since N is large and $\frac{q(x-\mu)}{N}$ is small we can expand the exponent and get

$$\int dx\, p(x) e^{\frac{iq(x-\mu)}{N}} = \int dx\, p(x) \left[1 + \frac{iq(x-\mu)}{N} - \frac{q^2(x-\mu)^2}{2N^2} + \ldots\right] = 1 - \frac{q^2\sigma^2}{2N^2} + \ldots$$

Returning to $\tilde{p}(z)$ we get (recall $\lim_{n\to\infty}\left(1+\frac{a}{n}\right)^n = e^a$)

$$\Rightarrow \tilde{p}(z) = \int \frac{dq}{2\pi} e^{iq(z-\mu)} \left[1 - \frac{q^2\sigma^2}{2N^2}\right]^N = \int \frac{dq}{2\pi} e^{iq(z-\mu)} e^{-\frac{q^2\sigma^2}{2N}} = \frac{1}{2\pi \frac{\sigma}{\sqrt{N}}} e^{-\frac{(z-\mu)^2}{2(\sigma/\sqrt{N})^2}}$$

which is a
normal distribution with mean μ (as expected) and variance $\sigma_N^2 = \frac{\sigma^2}{N}$.

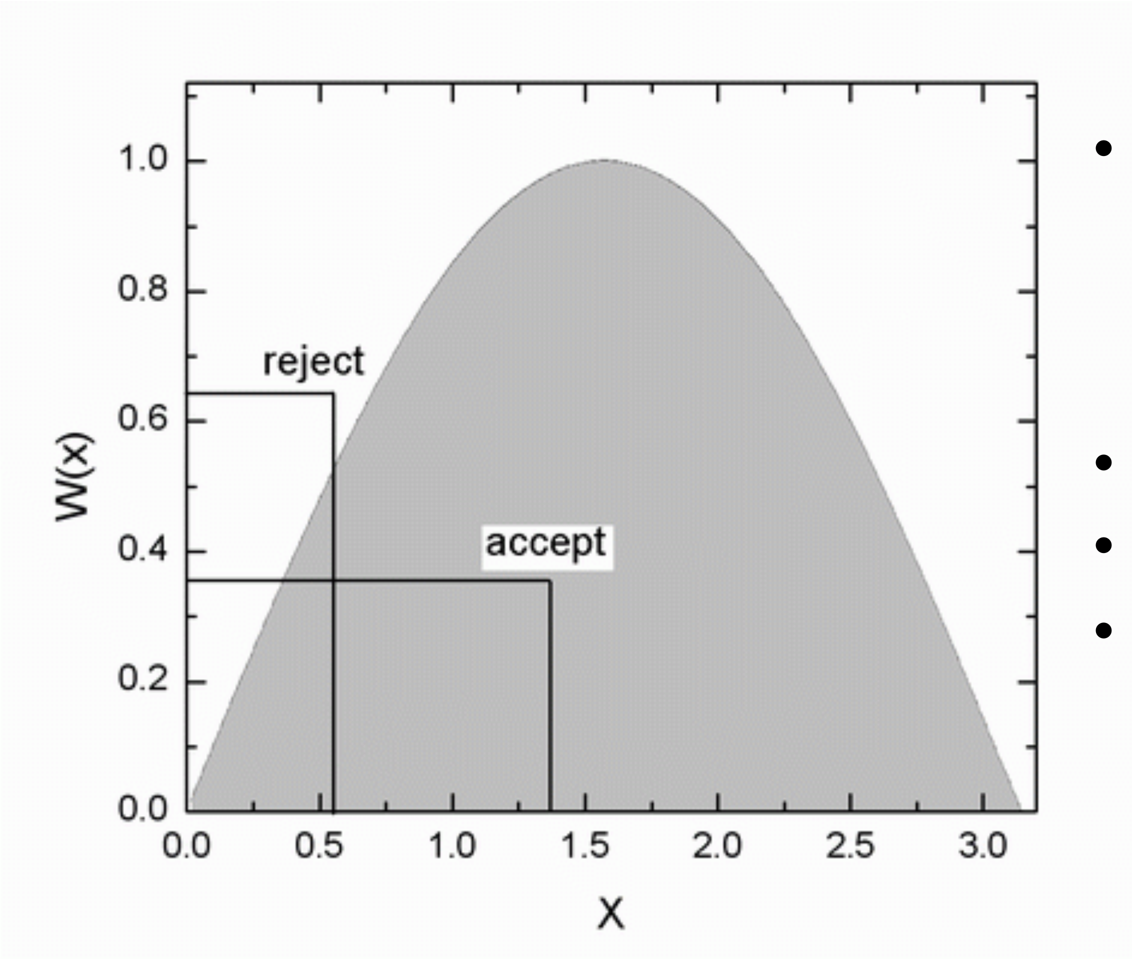
Note that the distribution $\tilde{p}(z)$ at large N is normal whatever (smooth) distribution $p(x)$ was.

From RAND to p(x)

- Random number generators give random numbers uniformly distributed in the interval [0,1].
- How to generate numbers distributed in an arbitrary interval [a,b] with PDF p(x)?

Method 1: von Neumann rejection

Generating non-uniform distribution with a probability distribution w(x)



- generate two random numbers
  x_i on $[x_{min}, x_{max}]$
  y_i on $[y_{min}, y_{max}]$
- if $y_i < w(x_i)$, accept
- if $y_i > w(x_i)$, reject
- The x_i so accepted will have the weighting w(x)

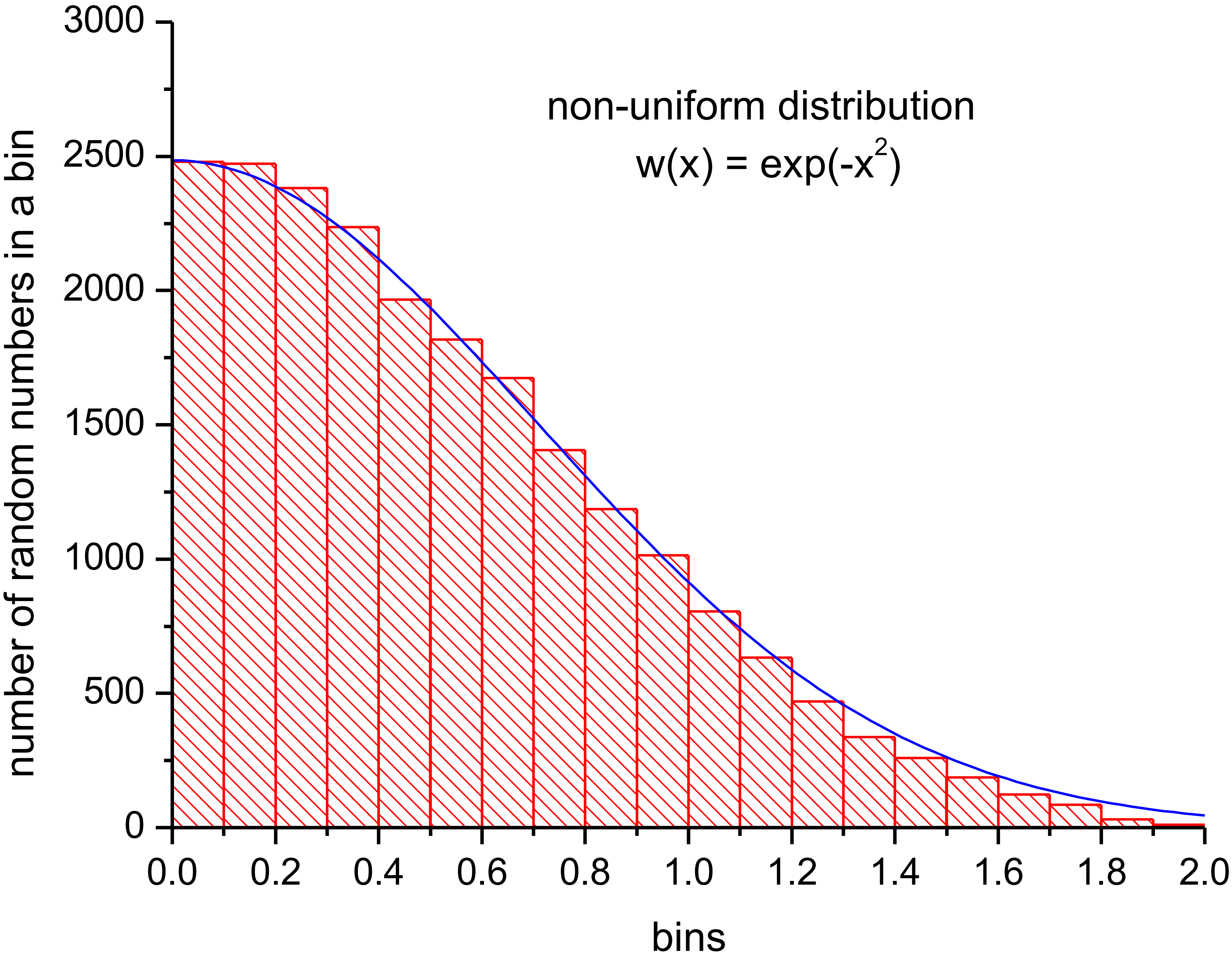
Example for $w(x)=exp(-x^2)$

```
double w(double);
int main ()
{
 int nmax = 50000;
 double xmin=0.0, xmax=2.0, ymin, ymax;
 double x, y;

 ymax = w(xmin);
 ymin = w(xmax);
 srand(time(NULL));
 for (double i=1; i <= nmax; i=i+1)
   {
      x = xmin + (xmax-xmin)*rand()/(RAND_MAX+1);
      y = ymin + (ymax-ymin)*rand()/(RAND_MAX+1);
      if (y > w(x)) continue;
      file_3  << " " << x << endl;   /* output to a file */
   }
 system("pause");
 return 0;
}

/* Probability distribution w(x) */
    double w(double x)
{
    return exp(0.0-1.0*x*x);
}
```

Example



3000
2500
2000
1500
1000
500
0
number of random numbers in a bin
non-uniform distribution
w(x) = exp(-x²)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
bins

Transformation of variables

The starting point is always the uniform distribution

$$p(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{everywhere else} \end{cases}$$

All random generators provide numbers distributed in this way.

When we attempt a transformation to a new variable $x \to y$ we have to conserve probability

$$p(x)dx = p(y)dy$$

For uniform distribution $p(x) = 1$ this implies

$$p(y)dy = dx$$

Transformation of variables

Suppose we have a $p(y)$ different from uniform PDF in the same interval [0,1]. Integrating the above equation we get the cumulative distribution of p(y):

$$x(y) = \int_0^y dy' \, p(y') \equiv P(y)$$

If we can invert this equation, transformation of variable

$$y = P^{-1}(x)$$

provides the non-uniform distribution with PDF $p(y)$.

Example 1

Suppose we have a general uniform PDF in the interval a,b].

$$p(y) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{everywhere else} \end{cases}$$

If we wish to relate this distribution to the one in the interval [0,1]

$$p(y)dy = \frac{dy}{b-a}$$

The cumulative function has the form

$$P(y) = \int_a^y \frac{dy}{b-a} = \frac{y-a}{b-a} = x(y)$$

and the inversion gives the anticipated result

$$y(x) = a + (b-a)x$$

From uniform to exponential

Assume that

$$p(y) = e^{-y}$$

is the exponential distribution in the interval $[0, \infty]$. The conservation of probability gives

$$p(y)dy = e^{-y}dy = dx$$

The cumulative

$$P(y) = \int_0^y dy' \, e^{-y'} = 1 - e^{-y} = x(y)$$

is easily inverted

$$y(x) = -\ln(1-x)$$

This gives us the new random variable y distributed with PDF e^{-y} in the interval $[0, \infty]$.

From uniform to normal

If a random number generator gives x distributed uniformly in the interval [0,1] the transformation

$$y(x) = -\ln(1-x)$$

gives us the new random variable y distributed with PDF e^{-y} in the interval $[0,\infty]$.

It may be implemented like that

$$\begin{aligned} &\ldots \\ x &= \text{rand}()/(\text{RAND\_MAX}+1) \\ y &= -\log(1-x) \\ &\ldots \end{aligned}$$

Example 3

Another example is the PDF

$$p(y) = \frac{(n-1)ba^{n-1}}{(a+by)^n}, \qquad \int_0^\infty dy\, p(y) = 1$$

(with $n > 1$) in the interval $[0, \infty]$.

The cumulative

$$P(y) = (n-1)ba^{n-1}\int_0^y dy' \frac{1}{(a+by')^n} = 1 - \frac{a^{n-1}}{(a+by)^{n-1}} = x(y$$

is easily inverted:

$$y(x) = \frac{a}{b}\left[(1-x)^{-\frac{1}{n-1}} - 1\right]$$

From uniform to normal

For the normal distribution $p(x)\ e^{-x^2}$ it is difficult to find an inverse since the cumulative is given by the error function erf(x).
However, for two normally distributed variables the trick to go to polar coordinates solves the problem.

Suppose x and y are two random variables distributed normally

$$g(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$$

Switch to polar coordinates

$$r = \sqrt{x^2+y^2}, \qquad \theta = \arctan\frac{x}{y}$$

results in

$$g(r,\theta) = \frac{1}{2\pi} e^{-r^2/2}$$

Normalization check

$$\frac{1}{2\pi}\int_0^\infty rdr \int_0^{2\pi} d\phi\ e^{-r^2/2} = \int_0^\infty rdr\ e^{-r^2/2} = 1$$

From uniform to normal

For PDF $e^{-r^2/2}$ the distribution over the angles is uniform so one can use the random number generator (rescaled to $[0, 2\pi]$ interval) whereas the distribution over r should be related to random numbers in the interval [0,1].

To do this, we introduce a new variable $u = \frac{r^2}{2}$ with the PDF $p(u) = e^{-u}$. From the results of Example 2 we see that it can be generated from random numbers $x' \ni [0, 1]$ by

$$u = -\ln(1 - x')$$

With

$$x = r\cos\theta = \sqrt{2u}\cos\theta$$
$$y = r\sin\theta = \sqrt{2u}\sin\theta$$

we can obtain new random numbers x, y in the interval $[0, \infty]$ by

$$x = \sqrt{-2\ln(1 - x')}\cos\theta$$
$$y = \sqrt{-2\ln(1 - x')|}\sin\theta$$

with x' in the interval [0,1] and θ in $[0, 2\pi]$.

More on integration – importance sampling

Importance sampling: more attention to regions corresponding to large values of the integrand

$$I = \int_a^b f(x)dx = \int_a^b \frac{f(x)}{p(x)} p(x)dx$$

where $p(x)$ is a probability density over x

The density $p(x)$ is called the importance function

Then with x_i from the distribution with density p

$$I \approx (b-a)\frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{p(x_i)}$$

Importance Sampling

Since random numbers are generated for the uniform distribution $p(x)$ with $x \in [0, 1]$, we need to perform a change of variables $x \to y$ through

$$x(y) = \int_a^y p(y')dy',$$

where we used

$$p(x)dx = dx = p(y)dy.$$

If we can invert $x(y)$, we find $y(x)$ as well.

Importance Sampling

With this change of variables we can express the integral of Eq. (61) as

$$I = \int_a^b p(y)\frac{F(y)}{p(y)}dy = \int_a^b \frac{F(y(x))}{p(y(x))}dx,$$

meaning that a Monte Carlo evalutaion of the above integral gives

$$\int_a^b \frac{F(y(x))}{p(y(x))}dx = \frac{1}{N}\sum_{i=1}^{N}\frac{F(y(x_i))}{p(y(x_i))}.$$

The advantage of such a change of variables in case $p(y)$ follows closely F is that the integrand becomes smooth and we can sample over relevant values for the integrand. It is however not trivial to find such a function p. The conditions on p which allow us to perform these transformations are

1. p is normalizable and positive definite,
2. it is analytically integrable and
3. the integral is invertible, allowing us thereby to express a new variable in terms of the old one.

Importance Sampling

The algorithm for this procedure is

- Use the uniform distribution to find the random variable y in the interval [0,1]. $p(x)$ is a user provided PDF.
- Evaluate thereafter

$$I = \int_a^b F(x)dx = \int_a^b p(x)\frac{F(x)}{p(x)}dx,$$

by rewriting

$$\int_a^b p(x)\frac{F(x)}{p(x)}dx = \int_a^b \frac{F(x(y))}{p(x(y))}dy,$$

since

$$\frac{dy}{dx} = p(x).$$

- Perform then a Monte Carlo sampling for

$$\int_a^b \frac{F(x(y))}{p(x(y))}dy, \approx \frac{1}{N}\sum_{i=1}^{N}\frac{F(x(y_i))}{p(x(y_i))},$$

with $y_i \in [0, 1]$,

- Evaluate the variance

Demonstration of Importance Sampling

$$I = \int_0^1 F(x)dx = \int_0^1 \frac{1}{1+x^2}dx = \frac{\pi}{4}.$$

We choose the following PDF (which follows closely the function to integrate)

$$p(x) = \frac{1}{3}(4-2x) \qquad \int_0^1 p(x)dx = 1,$$

resulting

$$\frac{F(0)}{p(0)} = \frac{F(1)}{p(1)} = \frac{3}{4}.$$

Check that it fullfils the requirements of a PDF. We perform then the change of variables (via the Cumulative function)

$$y(x) = \int_0^x p(x')dx' = \frac{1}{3}x(4-x),$$

or

$$x = 2 - (4-3y)^{1/2}$$

We have that when $y = 0$ then $x = 0$ and when $y = 1$ we have $x = 1$.

The Metropolis algorithm

In 1953 Metropolis introduced "the idea of importance sampling" that can considerably improve speed and quality of calculations.

$$I = \int_a^b w(x) f(x) dx \approx (b-a) \frac{1}{N} \sum_{i=1}^{N} w(x_i) f(x_i)$$

In the simplest version, $x_{i+1} = x_i + h(2u_i - 1)$ where *h* is a step and u_i is from a uniform random distribution

The step is accepted if

$$\frac{w(x_{i+1})}{w(x_i)} \geq \alpha_i$$

where α_i is a random number from a uniform distribution

Metropolis Algorithm

Another way of generating an arbitrary nonuniform probability distribution was introduced by Metropolis, Rosenbluth, and Teller in 1953. The Metropolis algorithm is a special case of an importance sampling procedure in which certain possible sampling attempts are rejected.

The Metropolis method is useful for computing averages of the form

$$\langle f \rangle = \frac{\int dx\, f(x) p(x)}{\int dx\, p(x)}$$

where $p(x)$ is an arbitrary probability distribution that need not be normalized.

Example: one-dimensional integral

For simplicity let us considerthe Metropolis algorithm for estimating one-dimensional definite integrals. Suppose that we wish to use importance sampling to generate random variables according to an arbitrary probability density $p(x)$. The Metropolis algorithm produces a random walk of points $\{x_i\}$ whose asymptotic probability distribution approaches $p(x)$ after a large number of steps. The random walk is defined by specifying a transition probability $T(x_i \to x_j)$ from one value x_i to another value x_j such that the distribution of points $x_0, x_1, ..., _i$ converges to $p(x)$.

It can be shown that it is sufficient (but not necessary) to satisfy the "detailed balance" condition

$$p(x_i)T(x_i \to x_j) \; = \; p(x_j)T(x_j \to x_i) \qquad (*)$$

The relation $(*)$ does not specify $T(x_i \to x_j)$ uniquely. A simple choice of $T(x_i \to x_j)$ that is consistent with Eq. $(*)$ is

$$T(x_i \to x_j) \; = \; \min\left(1, \frac{p(x_j)}{p(x_i)}\right)$$

Example

If the "walker" is at position x_i and we wish to generate x_{i+1}, we can implement this choice of $T(x_i \rightarrow x_j)$ by the following steps:

1. Choose a trial position $x_{\text{trial}} = x_i + \delta_i$ where δ_i is a random number in the interval $[-\delta, \delta]$.
2. Calculate $w = \frac{p(x_{\text{trial}})}{p(x_i)}$
3. If $w \geq 1$ accept the change and let $x_{i+1} = x_{\text{trial}}$.
4. If $w < 1$ generate a random number r.
5. If $r \leq w$, accept the change and $x_{i+1} = x_{\text{trial}}$.
6. If the trial change is not accepted, then let $x_{i+1} = x_i$

It is necessary to sample many points of the random walk before the asymptotic probability distribution $p(x)$ is attained.

How do we choose the maximum "stepsize" δ?

If δ is too large, only a small percentage of trial steps will be accepted and the sampling of $p(x)$ will be inefficient. On the other hand, if δ is too small, a large percentage of trial steps will be accepted, but again the sampling of $p(x)$ will be inefficient.

A rough criterion for the magnitude of δ is that approximately one third to one half of the trial steps should be accepted.

We also wish to choose the value of x_0 such that the distribution $\{x_i\}$ will approach the asymptotic distribution as quickly as possible. An obvious choice is to begin the random walk at a value of x at which $p(x)$ is a maximum.