Brief Introduction to Probability and Simulation: Part 1 — Probability, and Sampling

Elaine Spiller

Marquette University

TIFR Center for Applicable Mathematics: Monsoon Summer School

July 5, 2011

Suppose an event has a set of outcomes $E_1, E_2, \ldots$ and associated with each outcome is a probability $p_1, p_2, \ldots$ with $0 \leq p_k \leq 1$, where practically speaking

$$\begin{array}{lclll} p_k & = & 0 & \text{means} & E_k \text{ never occurs} \\ p_k & = & 1 & \text{means} & E_k \text{ always occurs} \end{array}$$

(note on notation $P(E_k) = p_k$)

Some basic probability rules:

- $P(E_i \text{ and } E_j) \leq p_i + p_j$, likewise $P(E_i \text{ or } E_j) \leq p_i + p_j$
- We say E_i and E_j are *mutually exclusive* if one happing means the other cannot.
  - $P(E_i \text{ and } E_j) = 0 \qquad P(E_i \text{ or } E_j) = p_i + p_j$
- $P(E_1 \text{ or } E_2 \text{ or } E_3 \text{ or } , \ldots) = \sum_i p_i = 1$

Compound experiments

A *compound experiment* is a set of experiments with its own set of outcomes and probabilities.

Consider a two experiment case with events $\{E_i\}$, $\{F_j\}$ and probabilities p_{1i}, p_{2j} respectively.

An outcome of this compound experiment is the pair (E_i, F_j) and p_{ij} is the probability of the compound event, called the *joint probability*.

More probability rules:

- Iff $p_{ij} = p_{1i}p_{2j}$, then E_i and F_j are *independent*.

More probability rules (continued)

- It's useful to rewrite the joint probability as

$$p_{ij} = \Big(\sum_k p_{ik}\Big)\frac{p_{ij}}{\sum_k p_{ik}} = p(i)\frac{p_{ij}}{\sum_k P_{ik}}.$$

  Here $p(i)$ defines the *marginal probability* that E_i occurs regardless of which F_j occurs. That is $p(i) = p_{1i} = \sum_k p_{ik}$.

- The *conditional probability* is

$$p(j|i) = \frac{p_{ij}}{\sum_k p_{ik}}$$

  the probability that F_j occurs given that E_i occurred.

Combining these two rules, we see

$$p_{ij} = p(i)p(j|i) = p(j)p(i|j)$$

Discrete random variables

If we associate a numerical value with each E_i, we get a *random variable*, X. The *expectation* of X is define as

$$E[X] = \sum_i p_i x_i$$

where $\{x_i\}$ are possible values that X could take on, we'll refer to this set as the *state space*. Note, that $E[X]$ is also referred to as the *mean* or *average value* of X.

Example: rolling a die

state space $= \{1, 2, 3, 4, 5, 6\}$

$p_i = 1/6$, so we see the expected value of a die role, X, is

$$\begin{aligned} E[X] &= \sum_i p_i x_i \\ &= \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = 3.5 \end{aligned}$$

More on expectations

Given a real-valued function, $g(x)$, and a random variable, X, then $g(X)$ is also a random variable and it's expectation is

$$E[g(X)] = \sum_i p_i g(x_i).$$

Example:

Event	p_i	x_i	$g(x) = 1 + 3x$
heads	1/2	1	4
tails	1/2	0	1

$E[X] = 1/2$ and

$E[g(X)] = 2.5$

Note, the expectation is a linear operator, so

$$E[\alpha g(X) + \beta h(X)] = \alpha E[g(X)] + \beta E[h(X)].$$

Higher moments and variance

The *n^th moment* of X is defined as

$$E[X^n] = \sum_i p_i x_i^n, \qquad E[x] = \sum_i p_i x_i = \mu = \text{mean}$$

The *n^th central moment* of X is defined as

$$E[(X-\mu)^n] = \sum_i p_i (x_i - \mu)^n$$

The *variance* of a random variable is the second central moment, or

$$\begin{aligned} var\{X\} = E[(X-\mu)^2] &= \sum p_i \big(x_i - E[X]\big)^2 \\ &= \ldots \text{some algebra} \ldots = E[X^2] - E^2[X] \end{aligned}$$

And likewise for $g(X)$, $var\{g\} = E[g^2] - E^2[g]$.

Independence

If $X, \ Y$ are random variable,

$$E[XY] = \sum_{i,j} p_{ij} x_i y_j$$

If $X, \ Y$ are <u>*independent*</u>, then $p_{ij} = p_{1i} p_{2j}$ and

$$E[XY] = \sum_{i,j} p_{1i} p_{2j} x_i y_j = \sum_{i} p_{1i} x_i \sum_{j} p_{2j} y_j = E[X]E[Y]$$

Covariance is a measure of independence

$$cov\{X, Y\} = E[XY] - E[X]E[Y]$$

If X and Y are independent, $cov\{X, Y\} = 0$ and

$$var\{X+Y\} = var\{X\}+var\{Y\}+2cov\{X, Y\} = var\{X\}+var\{Y\}$$

(note, converse is not generally true)

Continuous random variables

Let x be a real-valued. Define a *cumulative distribution function* (CDF) as

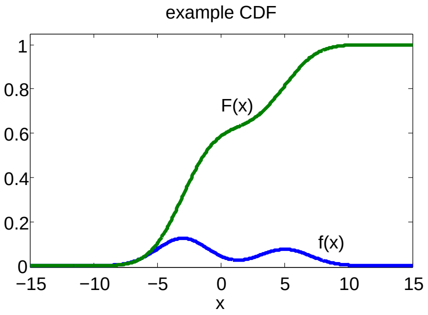
$$F(x) = P\{\text{a random variable } X < x\}$$

Some properties of CDFs:

- $P\{X < x_1\} + P\{x_1 \leq X < x_2\} = P\{X < x_2\}$
- $F(x)$ is a nondecreasing function, $0 \leq F(x) \leq 1$ for all x
- $F(-\infty) = 0$, $F(\infty) = 1$
- If $F(x)$ is differentiable, then the *probability density function* (pdf) is defined as

$$f(x) = \frac{dF}{dx}$$

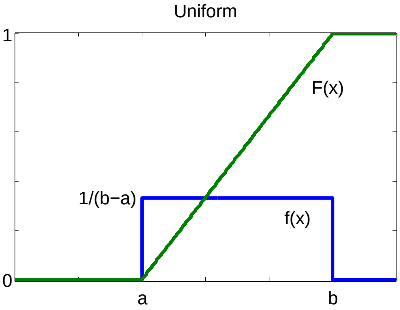
Example



Some properties and definitions of pdfs:

- $\int_{-\infty}^{\infty} f(x)dx = 1$ and $f(x) \geq 0$
- $P\{a \leq X < b\} = \int_a^b f(x)dx$
- mean = $E[X] = \mu = \int_{-\infty}^{\infty} xf(x)dx$
- variance = $\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$

Some useful distributions: Uniform

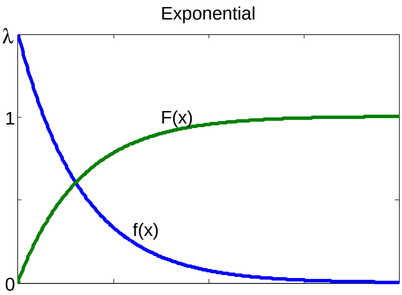


$$F(x) = \begin{cases} \frac{x-a}{b-a} & \text{for } a < x < b \\ 0 & \text{otherwise} \end{cases}$$

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a < x < b \\ 0 & \text{otherwise} \end{cases}$$

$$\text{mean} = E[X] = \frac{a+b}{2} \qquad \text{variance} = Var(X) = \frac{(b-a)^2}{12}$$

Some useful distributions: Exponential

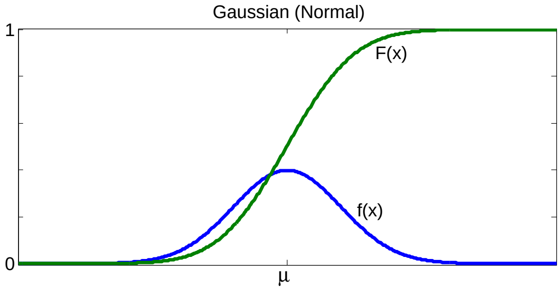


$$F(x) = 1 - \exp(-\lambda x) \qquad \text{for } 0 < x < \infty$$

$$f(x) = \lambda \exp(-\lambda x) \qquad \text{for } 0 < x < \infty$$

$$\text{mean } = E[X] = \frac{1}{\lambda} \qquad \text{variance } = Var(X) = \frac{1}{\lambda^2}$$

Some useful distributions: Gaussian

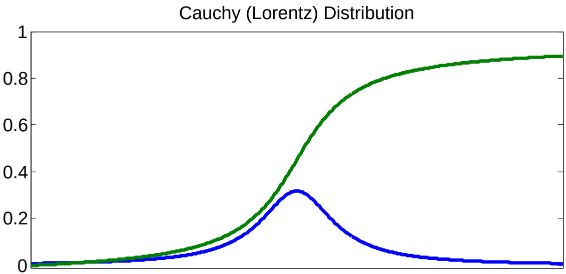


$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} \exp(-\frac{(t-\mu)^2}{2\sigma^2})dt$$

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

mean $= E[X] = \mu$ variance $= var(X) = \sigma^2$

Some useful distributions: Cauchy (Lorentz)



$$F(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(\frac{x}{a})$$

$$f(x) = \frac{1}{\pi}\frac{a}{a^2 + x^2}$$

mean $=$ doesn't exist variance $= \infty$

Joint Continuous Distributions

<u>Define</u> a *bivariate distribution function* and *joint probability density function*, respectively, as

$$F(x, y) = P\{X \leq x, Y \leq y\} \qquad f(x, y) = \frac{\partial^2 F}{\partial x \partial y}$$

And <u>define</u> the *covariance* matrix as

$$\Sigma = \begin{pmatrix} var(X) & cov\{X, Y\} \\ cov\{X, Y\} & var(Y) \end{pmatrix}$$

Where

$$cov\{X, Y\} = E[(X - \mu_x)(Y - \mu_y)] = E[XY] - E[X]E[Y]$$

$$cov\{X, Y\} = \iint xyf(x, y)dxdy - \left(\iint xf(x, y)dxdy \right) \left(\iint yf(x, y)dxdy \right)$$

Conditional probability and expectation

Note, If X, Y are *independent*, then $f(x, y) = f_1(x)f_2(y)$.

Conditional probability (probability of y, given x)

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

where $f_1(x) = \int f(x, y)dy$ is the *marginal probability* of x.

Conditional expectation (expectation of Y, given $X = x$)

$$E[Y|X = x] = \int yf(y|x)dy = \frac{\int yf(x, y)dy}{\int f(x, y)dy}$$

Note, $E[Y|X = x]$ is a function of x and

$$\begin{aligned} E[E[Y|X]] = \int \Big(\frac{\int yf(x, y)dy}{\int f(x, y)dy}\Big)f_1(x)dx &= \iint yf(x, y)dxdy \\ &= \int yf_2(y)dy = E[y] \end{aligned}$$

Bayes Theorem

Noting that

$$f(x) = \int f(x,y)dy \qquad \text{and} \qquad f(y) = \int f(x,y)dx,$$

and rewriting conditional probability from that last page, we see

$$f(x,y) = f(y|x)f(x), \qquad \text{and likewise} \qquad f(x,y) = f(x|y)f(y)$$

which tells us

$$f(x|y)f(y) = f(y|x)f(x) \qquad \text{or} \qquad f(x|y) = \frac{f(y|x)f(x)}{f(y)}.$$

Or as we saw yesterday,

$$f(x|y) \propto f(y|x)f(x)$$

How Bayes Theorem relates to data assimilation

What we're working with:

x – state variable
y – observation
$f(x)$ – prior distribution
$f(y|x)$ – likelihood of observation
$f(x|y)$ – posterior distribution, or how probable the state is given the observation

Goal: to know posterior distribution $f(x|y) \propto f(y|x)f(x)$

Options:

- Assume everything is Gaussian
- Draw many samples from $f(x|y)$ to approximate it

Random number generators: sampling pdfs

How can we use the uniform distribution to generate random variables from an arbitrary distribution?

(Assume we have access to $X \sim U(0,1)$, in MATLAB '*rand*'.)

Want:

- random numbers distributed according to (given) $q(y)$

Let:

- $p(x)$ be a uniform distribution.

Seek:

- a transformation $y = g(x)$, such that if $X \sim U(0,1)$, $Y \sim q$

Use:

- Fundamental Transformation Law

$$|q(y)dy| = |p(x)dx| \qquad \text{or} \qquad q(y) = p(x)\left|\frac{dx}{dy}\right|$$

Random number generators: integrate and invert

$$y = g(x) \quad \text{or} \quad x = g^{-1}(y) \quad \text{so} \quad \frac{dx}{dy} = \frac{d}{dy}[g^{-1}(y)]$$

From last page $q(y) = p(x)\frac{dx}{dy}$, but $p(x) = 1$ for $0 < x < 1$, so

$$q(y) = \frac{d}{dy}[g^{-1}(y)] = \frac{dx}{dy}$$

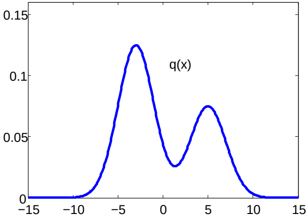
Now q (known) is the derivative of a function we'd like to know, so we integrate

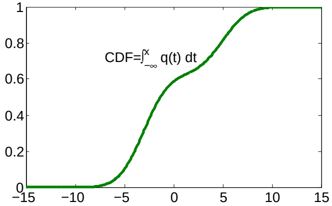
$$\begin{aligned} \int q(y)dy &= \int \frac{d}{dy}[g^{-1}(y)]dy = \int \frac{dx}{dy}dy \\ &= g^{-1}(y) = x \end{aligned}$$

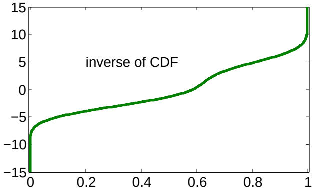
Now we have the inverse of the function we seek, so we invert

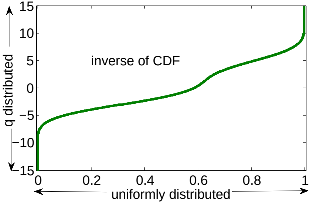
$$y = g(x)$$

Integrate and invert: pictorially









Difficulties with integrate and invert method

Difficulties:

- Sometimes can't integrate $q(y)$ analytically
- Sometimes can't invert $\int q(y)dy$
- This is a laborious/impossible task in multiple dimensions
- Sometimes (often) don't have analytic expression for $q(y)$. (Note, we might be able to evaluate q, but not write it down as a function)

But:

- Sampling distributions of interest is essential to many data assimilation schemes

First solution:

- Rejection sampling

Rejection sampling

Suppose we want to generate samples from $q(x)$ (and integrate/invert method is not feasible).

Choose:

- a function $f(x)$ such that $f(x) \geq q(x)$ for all x and
- $f(x)$ is a function that you can integrate and invert

Let:

- $A = \int_{-\infty}^{\infty} f(x)dx < \infty$ and $x = g(y) = F^{-1}(y)$ where
- $F(x) = \int_{-\infty}^{x} f(s)ds = y$

This construction of g enables us to generate random variables (rvs) with probability density $f(x)/A$.

Rejection sampling algorithm

- Select a uniform random variable Z in range of $(0, A)$
- Set $X = g(Z)$ (where g is constructed as on last page)
- Choose a uniform rv Y in range of $(0, f(X))$
- If $Y \leq q(X)$ accept X and $X \sim q$ or
- If $Y > q(X)$ reject X, and repeat process
- Note, $P(\text{acceptance}) = 1/A$