Probability Review

Giulia Iori & Gabriel Montes-Rojas[1]

CITY UNIVERSITY LONDON

[1]These slides are available at my webpage: http://www.staff.city.ac.uk/∼sbbc685

"How do you want it—the crystal mumbo-jumbo or statistical probability?"

Set theory

Sets

A **set** is a well-defined collection of objects. The objects of a set are called elements or members. The elements of a set can be anything: numbers, people, letters of the alphabet, other sets, and so on. Sets are conventionally denoted with capital letters, A, B, C, etc. Two sets A and B are said to be equal, written A = B, if they have the same members.

A set can also have zero members. Such a set is called the **empty set** (or the **null** set) and is denoted by the symbol ϕ.

If every member of the set A is also a member of the set B, then A is said to be a subset of B, written $A \subset B$, also pronounced A is contained in B.

Set operations

Union Let A and B be any two sets. The set which consists of all the points which are in A or B or both is defined the union and is written $A \cup B$.

Intersection Let A and B be any two sets. The set that consists of all the points which are both in A and B is defined the intersection and is written $A \cap B$ or AB.

Let $A_1, A_2, \cdots A_n \subset U$:
Disjoint if $A_i \cap A_j = \phi$ for any i, j.

Open sets, closed sets, and complements

A set U is called **open** if, intuitively speaking, you can "wiggle" or "change" any point x in U by a small amount in any direction and still be inside U. In other words, if x is surrounded only by elements of U; it can't be on the edge of U.

As a typical example, consider the open interval (0,1) consisting of all real numbers x with $0 < x < 1$. If you "wiggle" such an x a little bit (but not too much), then the wiggled version will still be a number between 0 and 1. Therefore, the interval (0,1) is open.

However, the interval (0,1] consisting of all numbers x with $0 < x \leq 1$ is not open; if you take $x = 1$ and wiggle a tiny bit in the positive direction, you will be outside of (0,1].

If $A \subset U$ then the complement of A in U, denoted by A^C, or $\bar{A}$ is:

$$A^C = U - A$$

A **closed** set is a set whose complement is open.

The real line, $\mathcal{R}$, is both an open and a closed set.

σ-algebra

A σ-algebra (or σ-field) over a set X is a family of subsets of X that is closed under countable set operations (union, intersection, complements); σ-algebras are mainly used in order to define measures on X. More formally,
Consider the collection of sets Ω. Then, $\mathcal{B}$ is a σ-algebra if:

1. $\Omega \in \mathcal{B}$
2. $B \in \mathcal{B}$ implies $B^C \in \mathcal{B}$.
3. $B_i, i \geq 1$ implies $\cup_{i=1}^{\infty} B_i \in \mathcal{B}$

A very important σ-algebra is the Borel σ-algebra. Consider $\mathcal{R}$, the real line and $\mathcal{C} = \{(a, b], -\infty \leq a \leq b \leq +\infty\}$, intervals that are open to the left and closed to the right.

Function

A **function** is a relation, such that each element of a set (the domain) is associated with a unique element of another (possibly the same) set (the codomain). We denote:

$$f : domain \rightarrow codomain$$

The most important functions for our purposes are the density and distribution functions, which we will introduce below.

Probability systems

We need to acquire an understanding of the different parts of a probability system and how they fit together. In order to make some sense of it all, we shall find it useful to think of a probability system as a physical experiment with a random outcome. To be more concrete, we shall use a specific example to guide us through the various definitions and what they signify.

Suppose that we toss a coin three times and record the results in order. This is a very simple experiment, but note that we should not necessarily assume that the coin toss is fair, with an equally likely outcome for heads or tails. There can, in principle, be many different probabilities associated with the same "physical experiment". This will have an impact on how we price derivatives.

Sample space

The basic entity in a probability system is the sample space, usually denoted Ω, which is a set containing all the possible outcomes of the experiment. If we denote heads by H and tails by T, then there are 8 different possible outcomes of the coin-tossing experiment, and they define the sample space as follows:

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

Definition:
The **sample space** $\Omega = \{\omega_i\}_{i=1}^{N}$ is the set of **all possible outcomes** of the experiment.

Event space

We are eventually going to want to talk about the probability of a specific "event" occurring. Is the sample space, simply as given, adequate to allow us to discuss such a concept? Unfortunately not.
This is because we want to ask more than just,

What is the probability that the outcome of the coin toss is a specific element of the sample space.

We also want to ask,

What is the probability that such-and-such specific events occur.

In order to be able to answer this, we need the concept of the set of all the events that we are interested in. This is called the event space, usually denoted Σ.

Event space

What conditions should an event space satisfy? The most "basic" event is Ω itself, that is, the event that one of the possible outcomes occurs. This event has probability one, that is, it always happens. It would thus make sense to require the event space to contain Ω.
Likewise, we shall assume that the null event ϕ, which occurs with probability zero, is also in the event space.
Next, suppose that the events $A = \{HTT, THH\}$ and $B = \{HTH, HHH, HTT\}$ are elements of Σ. It is natural to be interested in the event that either A or B occurs. This is the union of the events, $A \cup B = \{HTT, THH, HTH, HHH\}$. We would like Σ to be closed under the union of two of its elements.

Finally, if the event $C = \{HHH, HTH, HTT\}$ is an element of Σ, then the probability of it occurring is one minus the probability that the complementary event $\Omega - C = \{HHT, TTT, TTH, THT, THH\}$ occurs. Hence if an event is in Ω, we would also like its complement to be in Ω. We can summarise the definition of the event space as follows.

Event space

Definition: The **event space** Σ is a set of subsets of the sample space Ω, satisfying the following conditions:

1. $\Omega \in \Sigma$
2. if $A, B \in \Sigma$, then $A \cup B \in \Sigma$
3. if $A \in \Sigma$, then $\Omega - A \in \Sigma$

Note that for our purposes, we can take Σ to be the power set (the set of all subsets) of Ω. The power set of our example system is perhaps just slightly too large to comfortably write out. It contains $2^8 = 256$ elements.

The system consisting of the sample space and the event space (Ω, Σ) might appropriately be called a **possibility system**, as opposed to a **probability system** because all that it tells us are the possible outcomes of our experiment. It contains no information about how probable each event is. The so-called **probability measure** is an additional ingredient, that must be specified in addition to the pair (Ω, Σ).

Probability measure

Now suppose that we want to assign a probability to each event in Σ. We can do this by means of a probability measure $P : \Sigma \to [0, 1]$. For any event $A \in \Sigma$, P[A] is the probability that the event A occurs. Now, what conditions should we place on a probability measure? We have already constrained its values to lie between zero and one. Since the event Ω always occurs, its probability is one, i.e. $P[\Omega] = 1$. Finally, if we have two disjoint sets, then the probability of their union occurring should be equal to the sum of the probabilities of the disjoint sets. For example,

$$P[\{HHH, TTT\}] = Prob[\{HHH\}] + Prob[\{TTT\}] = \frac{1}{4}$$

Definition: A **probability measure** P is a function $P \rightarrow [0,1]$ satisfying
1. $P[A] \geq 0$ for every $A \in \Sigma$
2. $P[\Omega] = 1$
3. if $A, B \in \Omega$ and $A \cap B = \phi$, then $P[A \cup B] = P[A] + P[B]$.

Taken together, the sample space, event space and probability measure form a so-called probability system, denoted $\mathcal{P} = (\Omega, \Sigma, P)$.

Probability measure

We can in principle consider various probability measures on the same sample and event spaces (Ω, Σ).
In our coin tossing example, we have already considered the probability measure P that we obtain if the coin that we are tossing is fair. However, we could also define a probability measure $Q : \Sigma \rightarrow [0, 1]$ that is based on an unfair coin. Suppose that for the unfair coin we get heads with probability 1/3, and tails with probability 2/3.

Probability measure

Then the probability measure is defined by the probabilities
$Q(\{HHH\}) = 1/27$
$Q(\{HHT\}) = Q(\{HTH\}) = Q(\{THH\}) = 2/27$
$Q(\{HTT\}) = Q(\{TTH\}) = Q(\{THT\}) = 4/27$
$Q(\{TTT\}) = 8/27.$
Both measures are, in principle, valid to consider, so that when we are talking about probabilities related to the coin tossing, we must specify whether we are in the probability system $\mathcal{P} = (\Omega, \Sigma, P)$ or in the probability system $\mathcal{Q} = (\Omega, \Sigma, Q)$ or possibly in some other system based on another weighting of the coins.

Recap

Definition : Probability Space A probability space is the triplet $(\Omega, \Sigma, P[\cdot])$ where Ω is a sample space, Σ is the σ-algebra of events and $P[\cdot]$ is a probability function with domain Σ.

Definition : Probability Function A probability function is a function with domain Σ (the σ-algebra of events) and counterdomain the interval $[0, 1]$ which satisfy the following axioms:

- $P[A] \geq 0$ for every A in Σ
- $P[\Omega] = 1$
- If $A_1, A_2, ...$ is a sequence of mutually exclusive events in Σ than

$$P[\bigcup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} P[A_i]$$


I CAN GIVE YOU MY 93.4% ASSURANCE THAT THERE IS LESS THAN A 65.6% POSSIBILITY THAT THIS EXCERCISE WILL SIMPLY GENERATE 34.8% MORE MEANINGLESS STATISTICS
© Original Artist
Reproduction rights obtainable from
www.CartoonStock.com

Properties of $P[\cdot]$

Theorem: $P[\phi] = 0$
Theorem: If A is an event in Σ, than $P[\bar{A}] = 1 - P[A]$
Theorem: If A_1 and A_2 are events in Σ, than
$P[A_1] = P[A_1 \cap A_2] + P[A_1 \cap \bar{A}_2]$
Theorem (Law of Addition): If A_1 and A_2 are events in Σ, than
$P[A_1 \cup A_2] = P[A_1] + P[A_2] - P[A_1 \cap A_2]$
More generally for n events $A_1, A_2, \cdots, A_n$
$P[A_1 \cup \cdots \cup A_n] = \sum_{j=1}^{n} P[A_j] - \sum\sum_{i<j} P[A_i \cap A_j] + \sum\sum\sum_{i<j<k} P[A_i \cap A_j \cap A_k] + \cdots + (-1)^{n+1} P[A_1 \cap A_2 \cdots \cap A_n]$
If the events are mutually exclusive then $P[A_1 \cup \cdots \cup A_n] = \sum_{j=1}^{n} P[A_j]$
Theorem: If A_1 and A_2 are events is Σ and $A_1 \subset A_2$, than
$P[A_1] \leq P[A_2]$
Theorem (Boole's inequality): If $A_1, A_2, \cdots A_n$ are events in Σ then
$P[A_1 \cup \cdots \cup A_n] \leq P[A_1] + P[A_2] + \cdots + P[A_n]$

Conditional Probability and Independence

Let A and B be two events in Σ of the given probability space $(\Omega, \Sigma, P[\cdot])$.
The conditional probability of event A given B, denoted $P[A|B]$, is defined as

$$P[A|B] = \frac{P[A \cap B]}{P[B]}$$

if $P[B] > 0$ and is undefined is $P[B] = 0$.
In the following $P(AB)$ is sometimes used as a short notation for $P(A \cap B)$.

Conditional Probability and Independence

Theorem: If $A_1, A_2, \cdots, A_n$ are mutually exclusive than $P[A_1 \cup \cdots \cup A_n|B] = \sum_{i=1}^{n} P[A_i|B]$

Theorem : If A is an event in Σ, than $P[\bar{A}|B] = 1 - P[A|B]$

Theorem: If A_1 and A_2 are events in Σ, than $P[A_1 \cup A_2|B] = P[A_1|B] + P[A_2|B] - P[A_1 A_2|B]$

Theorem: If A_1 and A_2 are events is Σ and $A_1 \subset A_2$, than $P[A_1|B] \leq P[A_2|B]$

Theorem: If $A_1, A_2, \cdots A_n$ are events in Σ than $P[A_1 \cup \cdots \cup A_n|B] \leq \sum_{j=1}^{n} P[A_i|B]$

Conditional Probability and Independence

Theorem (Law of total probability): For a given probability space $(\Omega, \Sigma, P[\cdot])$, if $B_1, B_2, \cdots B_n$ is a collection of exhaustive and mutually disjoint events in Σ and $P[B_k] > 0$ for $k = 1, 2, \cdots n$, than for every A in Σ $P[A] = \sum_{j=1}^{n} P[A|B_j]P[B_j]$
Corollary: $P[A] = P[A|B]P[B] + P[A|\bar{B}]P[\bar{B}]$

Conditional Probability and Independence

Theorem (Bayes Formula): For a given probability space $(\Omega, \Sigma, P[\cdot])$, if $B_1, B_2, \cdots B_n$ is a collection of exhaustive and mutually disjoint events in Σ and $P[B_k] > 0$ for
$k = 1, 2, \cdots n$, than for every A in Σ s.t. $P[A] > 0$
$P[B_k|A] = \frac{P[A|B_k]P[B_k]}{\sum_{j=1}^{n} P[A|B_j]P[B_j]}$
Corollary: $P[B|A] = \frac{P[A|B]P[B]}{P[A]}$

Conditional Probability and Independence

Theorem (Law of Multiplication): For a given probability space $(\Omega, \Sigma, P[\cdot])$, let $A_1, A_2, \cdots A_n$ be events in Σ for which $P[A_1 \cdots A_n] > 0$, than

$$P[A_1 A_2 \cdots A_n] = P[A_1]P[A_2|A_1]P[A_3|A_1 A_2] \cdots P[A_n|A_1 \cdots A_{n-1}]$$

Corollary: $P[AB] = P[A|B]P[B]$

Conditional Probability and Independence

For a given probability space $(\Omega, \Sigma, P[\cdot])$, let A and B be two events in Σ. Events A and B are **independent** (or statistically independent or stochastically independent) if and only if any of the following conditions is satisfied:

- $P[A|B] = P[A]$ if $P[B] > 0$
- $P[B|A] = P[B]$ if $P[A] > 0$
- $P[AB] = P[A]P[B]$

Conditional Probability and Independence

Definition: Independence of several events For a given probability space $(\Omega, \Sigma, P[\cdot])$, let $A_1, A_2 \cdots A_n$ be n events in Σ. Events $A_1, A_2 \cdots A_n$ are defined to be independent if and only if

- $P[A_i A_j] = P[A_i]P[A_j]$ for $i \neq j$
- $P[A_i A_j A_k] = P[A_i]P[A_j]P[A_k]$ for $i \neq j, k \neq j, i \neq k$.

  $\vdots$
- $P[\cap_{i=1}^{n} A_i] = \prod_{i=1}^{n} P[A_i]$

Random variables

A **random variable** is a real-valued function X defined on the sample space Ω. Thus $X : \Omega \rightarrow \mathcal{R}$ assignes to each element ω_i of Ω an element of $\mathcal{R}$, that is, a real number. Even though the function is itself deterministic, that is, if we give X a definite input then we get a definite output, its argument ω_i is the random outcome of our physical experiment and hence $X(\omega_i)$ is also random.

Random variables

Consider the experiment of tossing 1 coin.
Example 1: $\Omega = \{\text{tail, head}\}$. We can define a random variable X so that $X(\omega) = 1$ if $\omega =$ head and $X(\omega) = 0$ if $\omega =$ tail.
Consider the experiment of tossing 3 coins.
Example2: X is be the function that counts the numbers of heads,
$X(\{HHH\}) = 3$
$X(\{HHT\}) = X(\{HTH\}) = X(\{THH\}) = 2$
$X(\{HTT\}) = X(\{TTH\}) = X(\{THT\}) = 1$
$X(\{TTT\}) = 0.$

Random variables

Definition: Random variable. For a given probability space $(\Omega, \Sigma, P[\cdot])$, a **random variable**, denoted by $X(\cdot)$, is a function with domain Ω and counterdomain the real line. The function $X(\cdot)$ makes some real number correspond to each outcome of the experiment.

Definition: distribution function. The (cumulative) distribution function a random variable X, denoted $F_X(\cdot)$, is that function with domain the real line and counterdomain the interval $[0, 1]$ which satisfies $F_X(x) = P[X \leq x] = P[\omega : X(\omega) \leq x]$ for every real number x.

- $F_X(-\infty) = 0$ and $F_X(+\infty) = 1$
- $F_X(\cdot)$ is a monotone non decreasing function, i.e. $F_X(a) \leq F_X(b)$ if $a < b$
- $F_X(\cdot)$ is continuous from the right, i.e. $\lim_{0<h\to 0} F_X(x + h) = F_X(x)$

[see graph]

Random variables

Definition: Discrete random variable A random variable will be defined discrete if the range of X is countable. If a random variable is discrete than the cumulative distribution function will be defined to be discrete.

Definition: Discrete density function If X is a discrete random variable with values $x_1, x_2, ...x_n, ...$ the function $f_X(x) = P[X = x_j]$ if $x = x_j, j = 1, ...n, ...$ and zero eitherwise is defined the discrete density function of X.

[see graph]

Random variables

Definition: Continuous random variable A random variable will be called continuous if there exists a function $f_X(\cdot)$ such that $F_X(x) = \int_{-\infty}^{x} f_X(u)du$ for every real number x.

Definition: Probability density function If X is a continuous random variable the function $f_X(\cdot)$ in $F_X(x) = \int_{-\infty}^{x} f_X(u)du$ is called the probability density function (or continuous density function). **[see graph]**

Note: It is important to recognize that f is not by itself a probability. Instead, the probability that X lies in the interval $(x, x + dx)$ is $f(x)dx$, or for a finite interval (a, b) it is $\int_a^b f(x)dx$.
Any function $f(\cdot)$ with domain the real line and counterdomain $[0, \infty)$ is defined to be a probability density function if:

- $f(x) \geq 0$ for all x
- $\int_{-\infty}^{\infty} f(x)dx = 1$

"Statisticians (and econometricians) do not know much, but they have their moments..."

Expectations and moments: Mean

Definition: Mean Let X be a random variable, the mean of X denoted by μ_X or $E[X]$ is defined by:

- $E[X] = \sum_j x_j f_X(x_j)$ if X is discrete with values $x_1, x_2, \cdots x_j, \cdots$. For the coin example: $E[X] = 0 \times P[X = 0] + 1 \times P[X = 1] = 0 \times 0.5 + 1 \times 0.5 = 0.5$.
- $E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$ if X is continuous.

Intuition: What value should we expect from X? If we have a considerable amount of draws from the X random variable, what would be their average?

- Let X and Y be two random variables, and c and d two constants. Then, $E[cX + dY] = cE[X] + dE[Y]$.
- Let X and Y be two **independent** random variables. Then, $E[X \times Y] = E[X] \times E[Y]$.

Expectations and moments: Variance

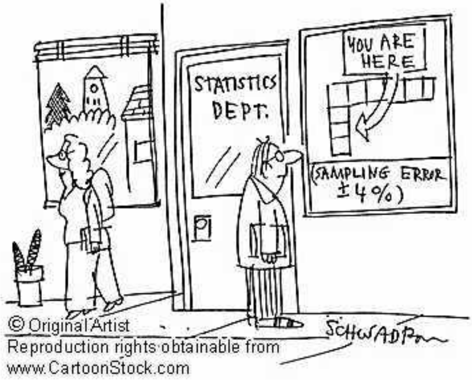
Definition: Variance Let X be a random variable, the variance of X denoted as σ_X^2 or $var[X]$ is defined by

- $var[X] = \sum_j (x_j - \mu_X)^2 f_X(x_j)$ id X is discrete with values $x_1, x_2, \cdots x_j, \cdots$.
- $var[X] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$ if X is continuous.

Intuition: However, for a given realisation of X, defined as x, we may have that $x \neq E[X]$. But, how much does it deviate from the E[X]?

- $var[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2$
- $var[aX] = a^2 \times var[X]$
- $var[aX + bY] = a^2 \times var[X] + b^2 \times var[Y] + ab \times cov[X, Y]$

Definition: Standard deviation Let X be a random variable, the standard deviation of X, denoted by σ_X is defined as $\sqrt{var[X]}$


YOU ARE HERE
STATISTICS DEPT.
(SAMPLING ERROR ±4%)
SCHWADRON
© Original Artist
Reproduction rights obtainable from
www.CartoonStock.com

Expectations and moments: Covariance

Idea: the **covariance** of the random variables X and Y measures how much co-movement they have.

Covariance and Correlation The covariance between two r.v. X and Y is defined as

$$cov(X,Y) = E[X - E[X]]E[Y - E[Y]] = E[XY] - E[X]E[Y]$$

Exercise: prove equality above.
The correlation coefficient is defined as

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}} = \frac{cov(X,Y)}{\sigma_(X)\sigma_(Y)}$$

Note that

- $-1 \leq \rho_{X,Y} \leq 1$
- $cov[X,X] = var[X]$ and $\rho_{X,X} = 1$
- If $cov[X,Y] = 0$ we say the two r.v. are uncorrelated.

Expectations and moments

Definitions: Moments Let X be a random variable, the r^{th} moments of X here denoted as μ_r is defined by $\mu_X^r = E[X^r]$
Definitions: Central moments Let X be a random variable, the r^{th} central moments of X here denoted as ν_X^r is defined by
$\nu_X^r = E[(X - \mu_X)^r]$
Hence the mean is the first moment and the variance the second central moment.

ν_X^3 is called skewness; symmetrical distribution have $\nu_X^3 = 0$.
μ_X^3/σ^3 is called coefficient of skeweness.
ν_X^4 is called kurtosis and measure the flatness of a density near its center.
$\nu_X^4/\sigma^4 - 3$ is called coefficient of kurtosis. A Normal density has a coefficient of kurtosis equal to zero.

Expectations and moments

Definitions: Moment generating function Let X be a random variable with density $f_X(\cdot)$. The moment generating function is defined as

$$m(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx$$

If a moment generating function exists than

$$\frac{d^r}{dt^r} m(t) = \int_{-\infty}^{\infty} x^r e^{tx} f_X(x)dx$$

Letting $t \to 0$ we have

$$\frac{d^r}{dt^r} m(0) = E[X^r] = \mu_X^r$$

A density function determines a set of moments (when they exist). But in general a sequence of moments does not determine a unique distribution function. However it can be proved that if a moment generating function exists than it uniquely determines the distribution function.

Expectations and moments

Definition: Expected value of a function of a random variable Let X be a random variable and $g(\cdot)$ a function with domain and counterdomain the real line. The expected value of $g(\cdot)$ denoted $E[g(X)]$, is defined

- $E[g(X)] = \sum_j g(x_j) f_X(x_j)$ if X is discrete with values $x_1, x_2, \cdots x_j, \cdots$.
- $E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$ if X is continuous.

Chebyshev Inequality

Suppose $\mathsf{E}(X^2)$ is finite, then

$$P[|X-\mu_X| \geq r\sigma_X] = P[(X-\mu_X)^2 \geq r^2\sigma_X^2] \leq \frac{1}{r^2}$$

Chebyshev Inequality

The proof is not difficult. Let $g(X)$ be a nonnegative function of the random variable X with domain the real line; than for any k

$$P[g(X) \geq k] \leq \frac{E[g(X)]}{k}$$

Proof:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx =$$

$$\int_{x:g(x)\geq k} g(x) f_X(x) dx + \int_{x:g(x)\leq k} g(x) f_X(x) dx$$

$$\geq \int_{x:g(x)\leq k} g(x) f_X(x) dx \geq \int_{x:g(x)\leq k} k f_X(x) dx$$

$$= kP[g(X) \geq k]$$

Now let $g(X) = (x - \mu_X)^2$ and $k = r^2\sigma_X^2$ and the Chebyshev inequality follows.

Chebyshev Inequality

Remark:

$$P[|X - \mu_X| \leq r\sigma_X] \geq 1 - \frac{1}{r^2}$$

so that

$$P[\mu_X - r\sigma_x < X < \mu_X + r\sigma_x] \geq 1 - \frac{1}{r^2}$$

that is, the probability that X falls within $r\sigma_X$ units of μ_X is greater than or equal to $1 - \frac{1}{r^2}$. For $r = 2$ one gets:

$$P[\mu_X - 2\sigma_x < X < \mu_X + 2\sigma_x] \geq \frac{3}{4}$$

or, for any random variable with finite variance at least three-fourth of the mass of X falls within two standard deviation from its mean. The Chebyshev inequality gives a bound, which does not depend on the distribution of X, for the probability of particular events in terms of a random variable and its mean and variance.

Jensen Inequality

Let X be a random variable with mean $E[x]$ and let $g(X)$ be a convex function; than $E[g(X)] \geq g([E[X])$. For example $g(x) = x^2$ is convex; hence $E[X^2] \geq (E[X])^2$, which guaranties that the variance of X is non-negative.

Important distributions

Discrete distributions

- Bernoulli
- Binomial
- Poisson

Continuous distributions

- Uniform
- Normal
- Exponential
- Gamma (also χ^2)
- t-Student (also Cauchy)

Discrete random variables

- Bernoulli distribution

$$domain(X) = \{0, 1\}$$

$$f_X(x) = p^x(1-p)^{(1-x)} I_{\{0,1\}}(x)$$

where $0 \leq p \leq 1$
(tossing a coin one time, head vs. tail)

- Binomial distribution

$$domain(X) = \{0, 1, ..., n\}$$

$$f_X(x) = \binom{n}{x} p^x(1-p)^{(n-x)} I_{\{0,1,2,\cdots,n\}}(x)$$

where $0 \leq p \leq 1$ and n is a positive integer. The binomial coefficient $\binom{n}{x} = \frac{n!}{(n-x)!x!}$
(tossing a coin n times and counting heads)

- Poisson

$$domain(X) = \{0, 1, 2, ...\}$$

$$f_X(x) = \frac{e^{-\lambda}\lambda^{-x}}{x!} I_{\{0,1,2,\cdots\}}(x)$$

where $\lambda > 0$
(how many pair of shoes do you own?)

Show that

- Bernoulli distribution

$$E[x] = p \qquad var[X] = pq \qquad m_x(t) = pe^t + q$$

  where q = 1-p

- Binomial distribution

$$E[x] = np \qquad var[X] = npq \qquad m_x(t) = (pe^t + q)^n$$

- Poisson

$$E[x] = \lambda \qquad var[X] = \lambda \qquad m_x(t) = e^{\lambda(e^t - 1)}$$

Uniform random variables

$$f_X(x) = \frac{1}{b-a} I_{\{a,x,b\}}$$

Show that

$$E[x] = \frac{a+b}{2} \qquad var[X] = \frac{(b-a)^2}{12} \qquad m_x(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

Normal random variables

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

Let's show that

$$I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$$

Change variable $y = \frac{(x-\mu)}{\sigma}$

$$I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

We do not know how to calculate I but we can calculate I^2

$$I^2 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

$$I^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\frac{-(y^2+z^2)}{2}} dydz$$

In polar coordinate $y = rcos\theta, \qquad z = rsin\theta, \qquad dydz = rd\theta dr$

$$I^2 = \frac{1}{2\pi}\int_0^{\infty}\int_0^{2\pi} e^{\frac{-r^2}{2}} rd\theta dr =$$

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

$$-\int_0^{\infty} de^{\frac{-r^2}{2}} = 1$$

Hence $I^2 = 1$ and also $I = 1$.

Exercise 1: Say X is $N(\mu, \sigma^2)$. Derive the moment generating function of X.

$$m_X(t) = E[e^{tX}] = e^{t\mu} E[e^{t(X-\mu)}] =$$

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

$$e^{t\mu} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}[(x-\mu)^2 - 2\sigma^2 t(x-\mu)]} dx =$$

$$e^{t\mu} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}[(x-\mu-\sigma^2 t)^2 - \sigma^4 t^2]} dx =$$

$$e^{t\mu + \frac{1}{2}\sigma^2 t^2} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}[(x-\mu-\sigma^2 t)^2]} dx = e^{t\mu + \frac{1}{2}\sigma^2 t^2}$$

From the moment generating function we can derive

$$m_X(t) = e^{t\mu + \frac{1}{2}\sigma^2 t^2}$$

we can derive

$$E[X] = m'_X(0) = \mu$$
$$var[X] = m''_X(0) - (m'_X(0))^2 = E[X^2] - E[X]^2 = \sigma^2$$

Exercise 2: Say X is $N(\mu, \sigma^2)$, where

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

Derive the p.d.f. of $W = e^X$.

Solution: We find the probability distribution function of W by first calculating its (cumulative) distribution function, that is,

$$F_W(w) = P(W \leq w) = P(e^X \leq w).$$

Since the exponential function is an increasing function, we can take its inverse on both sides of the last inequality to obtain

$$F_W(w) = P(X \leq \log w) = \int_{-\infty}^{\log w} \frac{1}{\sqrt{2\pi}\sigma} exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx, \qquad 0 < w.$$

Using Leibniz's rule for differentiating an integral

$$\frac{d}{dz}\int_{a(z)}^{b(z)} f(x,z)\, dx = \int_{a(z)}^{b(z)} \frac{\partial f(x,z)}{\partial z}\, dx + f(b(z),z)\frac{db(z)}{dz} - f(a(z),z)\frac{da(z)}{dz}$$

Therefore, the p.d.f for W is

$$f_W(w) = \frac{dF_W}{dw}(w) = \frac{1}{\sqrt{2\pi}\sigma w} exp\left[-\frac{(\log w - \mu)^2}{2\sigma^2}\right], \qquad 0 < w.$$

This distribution is called a *lognormal* distribution, for obvious reasons.


I CAN GIVE YOU MY 93.4% ASSURANCE THAT THERE IS LESS THAN A 65.6% POSSIBILITY THAT THIS EXCERCISE WILL SIMPLY GENERATE 34.8% MORE MEANINGLESS STATISTICS
© Original Artist
Reproduction rights obtainable from
www.CartoonStock.com

Exercise 3: Say $\log W \sim N(\mu, \sigma^2)$. Show that

$$E[W] = e^{\mu + \frac{1}{2}\sigma^2}$$

and

$$var[W] = e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2}$$

Solution:
Take $X \sim N(\mu, \sigma)$. Then $W = e^X$. Use

$$E[W] = E[e^X] = m_X(1)$$

and

$$varW = E[e^{2X}] - E[e^X]^2 = m_X(2) - m_X(1)^2$$

Exercise 4: Let $X \sim N(0,1)$. Derive the density function of $Z = X^2$, its moment generating function of and the first two moments.

<u>Solution:</u>

If X have a standard normal distribution $N(0,1)$ the distribution of their squares $Z = X^2$ is:

$$\begin{aligned} F_Z(z) = P(X^2 \leq z) &= P(-\sqrt{z} \leq X \leq \sqrt{z}) \\ &= \int_{-\sqrt{z}}^{\sqrt{z}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= 2\int_0^{\sqrt{z}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx. \end{aligned}$$

Using Leibniz's rule for differentiating an integral

$$\frac{d}{dz}\int_{a(z)}^{b(z)} f(x,z)\, dx = \int_{a(z)}^{b(z)} \frac{\partial f(x,z)}{\partial z}\, dx + f(b(z),z)\frac{db(z)}{dz} - f(a(z),z)\frac{db(z)}{dz}$$

the p.d.f for Z is

$$f_Z(z) = \frac{dF_Z}{dz} = \frac{1}{z\sqrt{2\pi}} e^{-z/2}.$$

This is the χ^2-distribution with parameter 1, or $\chi^2(1)$.

The moment generating function is

$$m_Z(t) = E[e^{tZ}] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tx^2} e^{-\frac{1}{2}x^2} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2(1-2t)} dx$$

Change variable $y = x(1-2t)^{\frac{1}{2}}$ and

$$m_Z(t) = (1-2t)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy = (1-2t)^{-\frac{1}{2}}$$

Hence $E[Z] = m'_Z(0) = 1$ and $var[Z] = m''_Z(0) - (m'_Z(0))^2 = 2$.

Exercise 5:
Let X_1 and X_2 be independent with normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$. Find the p.d.f. of $Y = X_1 - X_2$.
Solution: The moment generating function for Y is

$$
\begin{aligned}
m_Y(t) &= E(e^{t(X_1 - X_2)}) \\
&= E(e^{tX_1} e^{-tX_2}) \\
&= E(e^{tX_1}) E(e^{-tX_2}) \\
&= m_{X_1}(t) m_{-X_2}(t),
\end{aligned}
$$

since X_1 and X_2 are independent. Using the fact that X_1 and X_2 are normally distributed, we have that

$$m_{X_1}(t) = \exp\left(\mu_1 t + \frac{1}{2}\sigma_1^2 t^2\right)$$

and

$$m_{-X_2}(t) = m_{X_2}(-t) = \exp\left(-\mu_2 t + \frac{1}{2}\sigma_2^2 t^2\right).$$

Therefore

$$\begin{aligned} m_Y(t) &= \exp\left(\mu_1 t + \frac{1}{2}\sigma_1^2 t^2\right) \exp\left(-\mu_2 t + \frac{1}{2}\sigma_2^2 t^2\right) \\ &= \exp\left((\mu_1 - \mu_2)t + \frac{(\sigma_1^2 + \sigma_2^2)t^2}{2}\right). \end{aligned}$$

We recognise this as being the moment generating function of a normally distributed random variable with mean $\mu_1 - \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$. Thus the p.d.f. for Y is $N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$.
Exercise: Show that if $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ then $Y = X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Bivariate discrete distribution

If X and Y are random variables, then we can define a joint probability function $f(x, y) = P(X = x, Y = y)$ where $f(x, y)$ is non-negative and $\sum \sum f(x, y) = 1$. We then define the *marginal* distributions

$$f_X(x) = \sum_y f(x, y) \qquad f_Y(y) = \sum_x f(x, y)$$

These are the probability functions for X and Y separately.
The *conditional probability of X given* *Y* is:

$$P(X \,|\, Y) = \frac{P(XY)}{P(Y)}$$

Bivariate continuous distributions

In the same way, we can have continuous bivariate (or multivariate) probability distributions. For that, we have a probability density function $f(x, y)$ and again this isn't a probability on its own. Instead, the joint probability that X lies in the interval $(x, x + dx)$ and Y lies in the interval $(y, y + dy)$ is $f(x, y)dxdy$. Finite probabilities are then given by double integrals: $P[(X, Y) \in A] = \int\int f(x, y)dxdy$.

Any function $f(x, y)$ can be a bivariate probability function providing that it is non-negative and its integral over the entire $x - y$ plane is unity.

Bivariate continuous distributions

We can then define marginal densities:

$$f_X(x) = \int_{\infty}^{\infty} f(x,y)dy \qquad f_Y(y) = \int_{\infty}^{\infty} f(x,y)dx$$

The conditional densities are:

$$f(x \mid y) = \frac{f(x,y)}{f_Y(y)} \quad f(y \mid x) = \frac{f(x,y)}{f_X(x)}$$

The condition for the independence of X and Y is that

- The joint density factorizes

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

- the marginal densities are equal to the conditional densities:

$$f(x \mid y) = f_X(x) \qquad f(y \mid x) = f_Y(y)$$

Bivariate normal distribution

The bivariate normal distribution for two random variables X,Y, is given by

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y(1-\rho^2)^{1/2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 -2\rho\frac{x-\mu_X}{\sigma_X}\frac{y-\mu_Y}{\sigma_Y}+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}$$

where ρ is the correlation coefficient between X and Y. If $\rho = 0$

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y}\exp\left[-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right]\exp\left[-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]$$

$$= f_X(x)f_Y(y)$$

Multivariate normal distribution

The multivariate normal distribution for n random variables $\mathbf{X} = (X_1, X_2, \cdots X_n)$ is given by

$$f(x, y) = \frac{1}{(2\pi)^{n/2}(det\Sigma)^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mu)\Sigma^{-1}(\mathbf{x} - \mu)'\right\}$$

where $\mathbf{x}, \mu \in \mathcal{R}^n$, Σ is the variance-covariance matrix, Σ^{-1} its inverse and $det\Sigma$ its determinant.

Exercise 6:

Let $f_{X,Y}(x,y)$ be the normal bivariate distribution. Calculate
a) $\int_{-\infty}^{\infty} f(x,y)dxdy$
b) $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$
c) The moment generating function of the bivariate normal distribution $m_{X,Y}(t_1,t_2)$
d) The conditional distribution $f_{X|Y}(x|y)$

"How do you want it—the crystal mumbo-jumbo or statistical probability?"

Solution

(a) The bivariate normal distribution for two random variables X,Y, is given by

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y(1-\rho^2)^{1/2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 -2\rho\frac{x-\mu_X}{\sigma_X}\frac{y-\mu_Y}{\sigma_Y}+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}$$

To compute the integral of $f(x,y)$ over the entire plane, let us first do the change of variables

$$u = \frac{x-\mu_X}{\sigma_X} \qquad \text{and} \qquad v = \frac{y-\mu_Y}{\sigma_Y}.$$

Therefore,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-[1/2(1-\rho^2)](u^2-2\rho uv+v^2)}dudv}{2\pi(1-\rho^2)^{1/2}}.$$

Now we complete the squares for the variable u in the exponent as follows

$$\begin{aligned}\frac{1}{2(1-\rho^2)}(u^2-2\rho uv+v^2) &= \frac{1}{2(1-\rho^2)}(u^2-2\rho uv+\rho^2v^2-\rho^2v^2+v^2)\\ &= \frac{1}{2(1-\rho^2)}[(u-\rho v)^2+(1-\rho^2)v^2]\\ &= \frac{1}{2}\left[\left(\frac{u-\rho v}{\sqrt{1-\rho^2}}\right)^2+v^2\right].\end{aligned}$$

We are then led to the following change of variables

$$w=\frac{u-\rho v}{\sqrt{1-\rho^2}},$$

which reduces the integral to

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-w^2/2}dw\int_{-\infty}^{\infty} e^{-v^2/2}dv = 1.$$

(b) We now need to calculate the marginal distribution for the random variable X. This amounts to performing an integration only over the variable y, so we begin with the change of variables

$$v = \frac{y - \mu_Y}{\sigma_Y}.$$

This leads to

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x,y)dy = \frac{1}{2\pi\sigma_X(1-\rho^2)^{1/2}} \\ &\times \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)v + v^2\right]\right\} \\ &= \frac{\exp\left\{-\frac{(x-\mu_X)^2}{2(1-\rho^2)\sigma_X^2}\right\}}{2\pi\sigma_X(1-\rho^2)^{1/2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[v^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)v\right.\right. \end{aligned}$$

we complete the squares in the exponent of the integrand as follows

$$
\begin{aligned}
v^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)v &= v^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)v \\
&\quad +\rho^2\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - \rho^2\left(\frac{x-\mu_X}{\sigma_X}\right)^2 \\
&= \left(v - \rho\frac{x-\mu_X}{\sigma_X}\right)^2 - \rho^2\left(\frac{x-\mu_X}{\sigma_X}\right)^2.
\end{aligned}
$$

Back to the integral and with the change of variables

$$t = \frac{1}{(1-\rho^2)^{1/2}}\left(v - \rho\frac{x-\mu_X}{\sigma_X}\right)$$

we obtain

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f(x,y)dy \\
&= \frac{\exp\left\{\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right\}}{2\pi\sigma_X}\int_{-\infty}^{\infty} e^{-t^2/2}dt \\
&= \frac{1}{\sqrt{2\pi}\sigma_X}\exp\left\{\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right\}.
\end{aligned}
$$

(c) The joint moment generating function is defined as

$$\begin{aligned} m_{X,Y}(t_1,t_2) = m(t_1,t_2) &= E[e^{t_1X+t_2Y}] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1X+t_2Y} f(x,y)dxdy. \end{aligned}$$

We begin with the same change of variables as before, namely,

$$u = \frac{x-\mu_X}{\sigma_X} \qquad \text{and} \qquad v = \frac{y-\mu_Y}{\sigma_Y}.$$

This leads us to

$$\begin{aligned} m_{X,Y}(t_1,t_2) &= \frac{e^{t_1\mu_X+t_2\mu_Y}}{2\pi(1-\rho^2)^{1/2}} \\ &\times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1\sigma_X u+t_2\sigma_Y v-[1/2(1-\rho^2)](u^2-2\rho uv+v^2)} dudv. \end{aligned}$$

The total exponent in the integrand is

$$-\frac{1}{2(1-\rho^2)}[u^2-2(1-\rho^2)t_1\sigma_X u - 2\rho uv + v^2 - 2(1-\rho^2)t_2\sigma_Y v]$$

Completing the squares for the variable u we obtain

$$
\begin{aligned}
-\frac{1}{2(1-\rho^2)}\{[u &- (1-\rho^2)t_1\sigma_X - \rho v]^2 + v^2 - 2(1-\rho^2)t_2\sigma_Y v \\
&- (1-\rho^2)^2 t_1^2\sigma_X^2 - 2(1-\rho^2)t_1\sigma_X\rho v - \rho^2 v^2\},
\end{aligned}
$$

and completing the squares in the variable v we get

$$
\begin{aligned}
-\frac{1}{2(1-\rho^2)}\{[u &- (1-\rho^2)t_1\sigma_X - \rho v]^2 + (1-\rho^2)(v - t_2\sigma_Y - t_1\sigma_X\rho)^2 \\
&- (1-\rho^2)(t_1^2\sigma_X^2 + t_2^2\sigma_Y^2 + 2t_2\sigma_Y t_1\sigma_X\rho)\}.
\end{aligned}
$$

With the change of variables

$$
w = \frac{u - (1-\rho^2)t_1\sigma_X - \rho v}{\sqrt{1-\rho^2}} \quad \text{and} \quad z = v - t_2\sigma_Y - t_1\sigma_X\rho,
$$

the exponent becomes

$$
-\frac{1}{2}w^2 - \frac{1}{2}z^2 + \frac{1}{2}(t_1^2\sigma_X^2 + t_2^2\sigma_Y^2 + 2t_2\sigma_Y t_1\sigma_X\rho).
$$

Returning to the integral, we obtain

$$m_{X,Y}(t_1,t_2) = \frac{e^{t_1\mu_X + t_2\mu_Y\,\frac{1}{2}(t_1^2\sigma_X^2 + t_2^2\sigma_Y^2 + 2t_2\sigma_Y t_1\sigma_X\rho)}}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}w^2 - \frac{1}{2}z^2}dzdw$$

Therefore

$$m(t_1,t_2) = e^{t_1\mu_X + t_2\mu_Y\,\frac{1}{2}(t_1^2\sigma_X^2 + t_2^2\sigma_Y^2 + 2t_2\sigma_Y t_1\sigma_X\rho)}$$

(d) We obtain the conditional distributions from the joint and marginal distributions. The conditional distribution of X for a fixed value of Y is given by

$$
\begin{aligned}
f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} \\
&= \frac{\exp\left\{-\frac{1}{2\sigma_X^2(1-\rho^2)}\left[x-\mu_X-\frac{\rho\sigma_X}{\sigma_Y}(y-\mu_Y)\right]^2\right\}}{\sqrt{2\pi}\sigma_X(1-\rho^2)^{1/2}},
\end{aligned}
$$

where for $f_Y(y)$ we have used the formula obtained in **(b)** with x and y interchanged. As we can see, this is a normal distribution with mean $\mu_X + \frac{\rho\sigma_X}{\sigma_Y}(y-\mu_Y)$ and variance $\sigma_X^2(1-\rho^2)$.

Joint density of functions of random variables

Assume you know the joint density function of two r.v. X_1, X_2 and you want to calculate the joint density of $Y_1 = g_1(X_1, X_2), Y_2 = g(X_1, X_2)$. One can show that

$$f_{Y_1,Y_2} = |J| f_{X_1,X_2}(g_1^{-1}(y_1, y_2), g_2^{-1}(y_1, y_2) I_D(y_1, y_2)$$

where J is the Jacobian i.e. the determinant of the matrix M whose elements are $M_{i,j} = \frac{\partial x_i}{\partial y_j}$ and $|J|$ is its absolute value.

Box-Muller approach

This is a method for generating normally distributed random variables starting from uniformly distributed random variables. Let $X_1 \sim U(0,1), X_2 \sim U(0,1)$

$$y_1 = \sqrt{-2\log x_1} cos(2\pi x_2)$$

$$y_2 = \sqrt{-2\log x_1} sin(2\pi x_2)$$

Then

$$x_1 = \exp\left(-\frac{y_1^2 + y_2^2}{2}\right)$$

$$x_2 = \frac{1}{2\pi} arctg(y_2/y_1)$$

Now

$$\frac{\partial x_1}{\partial y_1} = -y_1 \exp\left(-\frac{y_1^2 + y_2^2}{2}\right)$$

$$\frac{\partial x_1}{\partial y_2} = -y_2 \exp\left(-\frac{y_1^2 + y_2^2}{2}\right)$$

$$\frac{\partial x_2}{\partial y_1} = -\frac{y_2}{2\pi} \frac{1}{y_1^2 + y_2^2}$$

$$\frac{\partial x_2}{\partial y_2} = -\frac{y_1}{2\pi} \frac{1}{y_1^2 + y_2^2}$$

and

$$|J| = \frac{1}{2\pi} \exp\left(-\frac{y_1^2 + y_2^2}{2}\right)$$

$$D = \{-\infty < y_1 < \infty, -\infty < y_2 < \infty\}$$

$$f_{Y_1,Y_2} = \frac{1}{2\pi} \exp\left(-\frac{y_1^2 + y_2^2}{2}\right) I_D(y_1, y_2)$$

Hence Y_1, Y_2 are independent normally distributed random variables. Now

$$Z_1 = Y_1$$

and

$$Z_2 = \rho Y_1 + \sqrt{1-\rho^2} Y_2$$

are normally distributed and ρ correlated.

Limit theory

The strong law of large numbers Let $X_1, X_2, \cdots X_n$ be a sequence if i.i.d r.v. each having a finite mean $\mu = E[X_i]$. Then

$$P\left\{\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n} = \mu\right\} = 1$$

Central Limit Theory Let $X_1, X_2, \cdots X_n$ be a sequence if i.i.d r.v. each having a finite mean μ and variance σ^2. Then

$$\lim_{n\to\infty} P\left\{\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{\frac{-x^2}{2}} dx = N(a)$$

Stochastic processes

A stochastic process is a sequence of a r.v. $X = (X_t(\omega), t \in T, \omega \in \Omega)$ defined on some probability space Ω. T can be a finite set, or countably infinite set (discrete-time process) or an interval (a, b) or (a, ∞) (continuous-time process). The index t is usually refereed to as time. A stochastic process $X_t(\omega)$ is a function of two variables:

- at fixed time t, $X_t(\omega), \omega \in \Omega$ it is a random variable
- for a given random outcome ω, $X_t(\omega), t \in T$ it is a function of time, called a realization or a trajectory, or sample path of the process X.

Random walk

A random walk is a formalization of the intuitive idea of taking successive steps, each in a random direction. A random walk is a simple stochastic process sometimes called a "drunkard's walk".
To generate a random walk

$$W(t+\Delta t) = W(t) + \sigma\epsilon\sqrt{\Delta t}$$

where $\epsilon \sim N(0,1)$.

Wiener process or Brownian motion

As the step size Δt in the random walk tends to 0 (and the number of steps increased comparatively) the random walk converges to Brownian motion in an appropriate sense.
A stochastic process $W = (W(t) : t \geq 0)$ is a standard Brownian if

- $W(0) = 0$
- W has independent increments: $W(t+u) - W(t)$ is independent of $(W(s) : s \leq t)$ for $u \geq 0$.
- W has stationary increments: the law of $W(t+u) - W(t)$ depends only on u.
- W has Gaussian increments: $W(t+u) - W(t) \sim N(0,u)$.