1.1 Review of Probability Theory

Angela Peace

Biomathemtics II
MATH 5355 Spring 2017

Lecture notes follow: Allen, Linda JS. An introduction to stochastic processes with applications to biology. CRC Press, 2010.

Basic Concepts and Definitions

Sample Space

Set S collection of elements

σ-algebra

A collection of events (subsets in S) $\mathcal{A}$ is a σ-algebra if

1. $S \in \mathcal{A}$
2. If $B \in \mathcal{A}$, then the complement of B is in $\mathcal{A}$. i.e. $B^c = \{s : s \in S, S \notin B\} \in \mathcal{A}$
3. If sequence $\{B_n\}_{n=1}^{\infty} \in \mathcal{A}$ then the union $\bigcup_{n=1}^{\infty} B_n \in \mathcal{A}$

Here $\mathcal{A}$ is **measurable** and the ordered pair $(S, \mathcal{A})$ is a **measurable space**.

Basic Concepts and Definitions

Probability Measure

P, a real-valued function defined on σ-algebra $\mathcal{A}$. The set function $P : \mathcal{A} \to [0, 1]$ is a probability measure if

1. $P(B) \geq 0$ for all $B \in \mathcal{A}$
2. $P(S) = 1$
3. If $B_i \bigcap B_j = \emptyset$ for $j, i = 1, 2, ..., i \neq j$ (pairwise disjoint) then $P(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} P(B_i)$ where $B_i \in \mathcal{A}$ for $i = 1, 2, ...$

Basic Concepts and Definitions

Probability Space

ordered triple $(S, \mathcal{A}, P)$

- S is the sample space
- $\mathcal{A}$ is the collection of events (subsets in S) and is a σ-algebra:
  1. $S \in \mathcal{A}$
  2. If $B \in \mathcal{A}$, then the complement of B is in $\mathcal{A}$. i.e. $B^c = \{s : s \in S, S \notin B\} \in \mathcal{A}$
  3. If sequence $\{B_n\}_{n=1}^{\infty} \in \mathcal{A}$ then the union $\bigcup_{n=1}^{\infty} B_n \in \mathcal{A}$
- P is a probability measure:
  1. $P(B) \geq 0$ for all $B \in \mathcal{A}$
  2. $P(S) = 1$
  3. If $B_i \bigcap B_j = \emptyset$ for $j, i = 1, 2, ..., i \neq j$ (pairwise disjoint) then $P(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} P(B_i)$ where $B_i \in \mathcal{A}$ for $i = 1, 2, ...$

Basic Concepts and Definitions

Conditional Probability

For events B_1, $B_2 \in \mathcal{A}$. The conditional probability of event B_1 given event B_2 is

$$P(B_1|B_2) = \frac{P(B_1 \cap B_2)}{P(B_2)}.$$

The conditional probability of event B_2 given event B_1 is

$$P(B_1|B_2) = \frac{P(B_1 \cap B_2)}{P(B_1)}.$$

Independent Events

B_1 and B_2 are independent if and only if

$$P(B_1 \cap B_2) = P(B_1)P(B_2)$$

i.e. $P(B_1|B_2) = P(B_1)$ and $P(B_2|B_1) = P(B_2)$.

Basic Concepts and Definitions

Random Variable

$(S, \mathcal{A})$ measurable space. A random variable X is a real function on S. $X : S \to \mathbb{R}$ such that

$$X^{-1}(-\infty, a] = \{s, X(s) \leq a\} \in \mathcal{A}.$$

Let A be the range of X. $A = \{x, X(s) = x, s \in S\}$. A is the **state space** of X.

- If A is finite or countable infinite then X is a **discrete random variable**
- If A is an interval (finite or infinite in length) then X is a **continuous random variable**

Example

Two sequential fair coin tosses

- What is the sample space?
  - $S = \{HH, HT, TH, TT\}$
- X is the discrete random variable with state space $A = \{1, 2, 3, 4\}$:
  - $X(HH) = 1$, $X(HT) = 2$, $X(TH) = 3$, $X(TT) = 4$
- Each event has probability $1/4$:
  - $P(\{HH\}) = \frac{1}{4}$, $P(\{HT\}) = \frac{1}{4}$, $P(\{TH\}) = \frac{1}{4}$, $P(\{TT\}) = \frac{1}{4}$
- Let B_1 be the event that 1^{st} coin toss is head and B_2 the event that the 2^{nd} coin toss is head:
  - $B_1 = \{HH, HT\} = \text{set } \{1, 2\}$
  - $B_2 = \{HH, TH\} = \text{set } \{1, 3\}$
- We can see B_1 and B_2 are **independent** events:

$$P(B_2|B_1) = \frac{P(B_1 \cap B_2)}{P(B_1)} = \frac{P(\{HH\})}{P(B_1)} = \frac{1/4}{1/2} = \frac{1}{2} = P(B_2)$$

Basic Concepts and Definitions

Notation

- Since random variable is defined as $X : S \to \mathbb{R}$. Events can be related to subsets of $\mathbb{R}$.
  - Ex: $B_1 = \{HH, HT\} =$ set $\{1, 2\}$
- $X = x$ is shorthand for the events $\{s : X(s) = x, s \in S\}$
- $X \leq x$ is shorthand for the events $\{s : X(s) \leq x, s \in S\}$
- The probability measure can be defined on $\mathbb{R}$:

$$P_X : \mathbb{R} \to [0, 1]$$

Basic Concepts and Definitions

Cumulative Distribution Function

c.d.f. of the random variable X is the function $F : \mathbb{R} \to [0, 1]$ defined by

$$F(x) = P_X((-\infty, x]).$$

F is nondecreasing, right continuous, and satisfies

$$\lim_{x \to \infty} F(x) = F(-\infty) = 0 \qquad \& \qquad \lim_{x \to \infty} F(x) = 1.$$

- describes how the probabilities accumulate

Basic Concepts and Definitions

Probability measure for *discrete* random variable

Probability Mass Function

For discrete random variable X, the function $f(x) = P_X(X = x)$ is the p.m.f. of X.

Some properties of f:

$$\sum_{x \in A} f(x) = 1 \qquad \& \qquad P_X(X \in B) = \sum_{x \in B} f(x)$$

for any $B \subset A$. The c.d.f. F of a discrete random variable satisfies

$$F(x) = \sum_{a_1 \leq x} f(a_i)$$

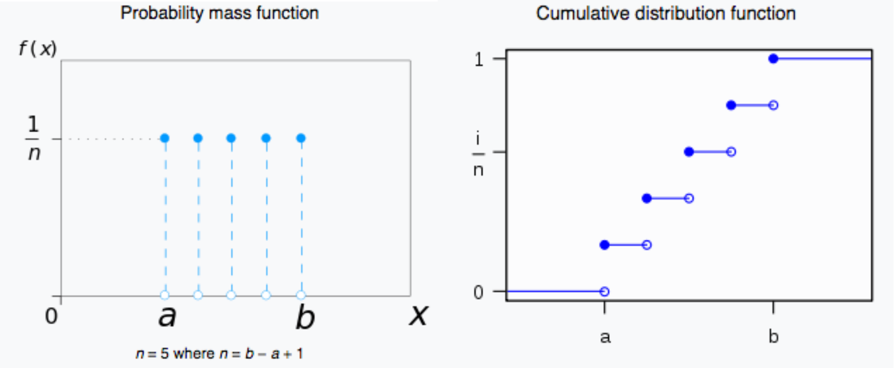
where a_i are the elements of A.

Example

Discrete uniform distribution

A finite number of values are equally likely to be observed; every one of n values has equal probability 1/n.

Let $A = \{1, 2, 3, 4, 5\}$ and $f(x) = 1/5$ for $x \in A$.



Basic Concepts and Definitions

Probability measure for *continuous* random variable

Probability Density Function

For continuous random variable X with c.d.f F. If there exists a nonnegative, integrable function $f : \mathbb{R} \to [0, \infty)$:

$$F(x) = \int_{-\infty}^{x} f(y)dy$$

then f is the p.d.f of X.

Some properties of f:

$$P_X(X \in A) = \int_A f(x)dx = 1 \qquad \& \qquad P_X(X \in B) = \int_B f(x)dx$$

for any $B \subset A$, In particular:

$$P_X(a \leq X \leq b) = \int_a^b f(x)dx = F(b) - F(a)$$

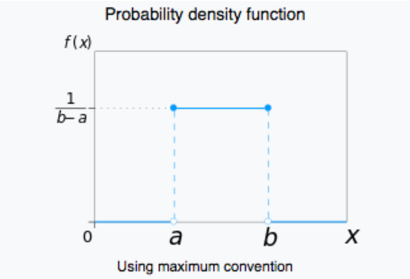
Note for a continuous random variable $P_X(a < X < b) = P_X(a \leq x \leq b)$.

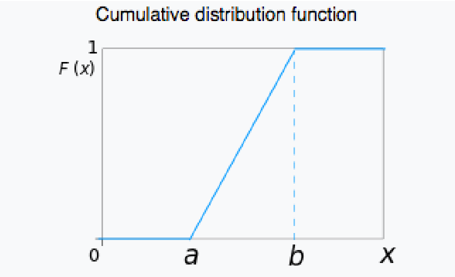
Example

Continuous uniform distribution

Each value is equally likely to be observed. For p.d.f.:

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





Probability Distributions

Discrete Distributions

Discrete Uniform

$$f(x) = \begin{cases} \frac{1}{n} & \text{for } x = 1, 2, ..., n \\ 0 & \text{otherwise} \end{cases}$$

Geometric

$$f(x) = p(1-p)^x \text{ for } x = 0, 1, ... \text{ and } 0 < p < 1$$

Here, p is the portability of success and $f(x)$ is the probability of one success in $x + 1$ trials

Probability Distributions

Discrete Distributions

Binomial

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{for } x = 0, 1, 2, ..., n \\ 0 & \text{otherwise} \end{cases}$$

where n is a positive integer, $0 < p < 1$ is the probability of success, and

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}.$$

It is denoted as $b(n, p)$. Think of $f(x)$ as the probability of x successes in n trials.

We will see this is the solution for a simple death process.

Probability Distributions

Discrete Distributions

Negative Binomial

$$f(x) = \begin{cases} \binom{x+n-1}{n-1} p^n (1-p)^x & \text{for } x = 0, 1, 2, \dots \\ 0 & \text{otherwise} \end{cases}$$

where n is a positive integer and $0 < p < 1$ is the probability of success.
Think of $f(x)$ as the probability of n successes in $n + x$ trials.
For $n = 1$ it simplifies to the geometric distribution.

We will see this is the solution for a simple birth process.

Probability Distributions

Discrete Distributions

Poisson

$$f(x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!} & \text{for } x = 0, 1, 2, ... \\ 0 & \text{otherwise} \end{cases}$$

where λ is a positive constant.

This distribution is important in continuous-time Markov chain models.

Probability Distributions

Continuous Distributions

Uniform

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

where $a < b$ are constants. It is denoted as $U(a, b)$.

The uniform distribution is the basis for random number generators, which are used for many stochastic models.

Probability Distributions

Continuous Distributions

Gamma

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

where α, β are positive constants, and $\Gamma(\alpha) = \int_0^\infty \frac{1}{\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx$. For a positive integer n, $\Gamma(n) = (n-1)!$.

Exponential

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

where λ is the positive constant.

Exponential d. is a special case of the Gamma d. with $\alpha = 1, \beta = 1/\lambda$. These distributions are associated with waiting time distributions.

Probability Distributions

Continuous Distributions

Normal

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

for $-\infty < x < \infty$, where μ, σ are constants.

- A random variable X that is uniformly distributed with mean μ and variance σ^2 is denoted as $X \sim N(\mu, \sigma^2)$.
- The standard normal distribution is $N(0, 1)$.
- The normal distribution is the underlying distribution for brownian motion (diffusion process).

Expectation

Expectation of X

For continuous random variable X with p.d.f. f, the expectation is defined as

$$E(X) = \int_{\mathbb{R}} xf(x)dx.$$

For discrete random variable X with probability function f defined on space $A = \{a_i\}_{i=1}^{\infty}$, the expectation is defined as

$$E(X) = \sum_{i=1}^{\infty} a_i f(ai).$$

- helps characterize the p.d.f. of a random variable.
- $E(X)$ is a weight average: the p.d.f. f is weighted by the values of the random variable X.

Expectation

Properties

- Expectations of a function of a random variable:

$$E(u(X)) = \int_{\mathbb{R}} u(x)f(x)dx$$

$$E(u(X)) = \sum_{i=1}^{\infty} u(a_i)f(ai)$$

- The expectation is a linear operator defined on set of functions $u(X)$:

$$E(a_1 u_1(X) + a_2 u_2(X)) = a1E(u_1(X)) + a2E(u_2(X))$$

- $E(b) = b$ for constant b
- We define the mean, variance, and moments of X in terms of the expectation.

Mean, Variance, Moments

Mean

The mean of the random variable X is $\mu_X = E(X)$.

Variance and Standard Deviation

The variance of the random variable X is $\sigma_X^2 = E([X - \mu_X]^2)$.
The standard deviation of X is σ.
Notation: $\sigma^2 = \sigma_X^2 = \text{VAR}(X)$.

Moments

The nth moment of X about point a is $E([X - a]^n)$.

The mean is the 1^{st} moment about the origin.
The variance is the 2^{nd} moment about the mean.

$$\sigma_X^2 = E([X - \mu_X]^2) = E(X^2) - 2\mu_X E(X) + \mu_X^2 = E(X^2) - \mu_X^2$$

Example

Discrete uniform distribution

X is a random variable with a discrete uniform distribution.

1. What is the mean of X?

$$\mu_X = E(X) = \sum_{x=1}^{n} x f(x) = \sum_{x=1}^{n} x \frac{1}{n} = \frac{1}{n} \sum_{x=1}^{n} x = \frac{1}{n} \left(\frac{n(n+1)}{2} \right) = \frac{n+1}{2}$$

2. What is the variance of X? $\qquad \sigma_X^2 = E(X^2) - \mu_X^2$

$$E(X^2) = \sum_{x=1}^{n} x^2 \frac{1}{n} = \frac{1}{n} \sum_{x=1}^{n} x^2 = \frac{1}{n} \left(\frac{n(n+1)(2n+1)}{6} \right) = \frac{(n+1)(2n+1)}{6}$$

$$\sigma_X^2 = E(X^2) - \mu_X^2 = \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2} \right)^2 = \frac{n^2-1}{12}$$

Example

Continuous uniform distribution

Y is a random variable with a discrete uniform distribution. Y is distributed as $U(0,1)$.

1. What is the mean of Y?

$$\mu_Y = E(Y) = \int_0^1 yf(y)dy = \int_0^1 y\frac{1}{1-0}dy = \int_0^1 y\ dy = \frac{1}{2}$$

2. What is the variance of Y? $\sigma_Y^2 = E(Y^2) - \mu_Y^2$

$$E(Y^2) = \int_0^1 y^2 f(y)dy = \int_0^1 y^2\ dy = \frac{1}{3}$$

$$\sigma_Y^2 = E(Y^2) - \mu_Y^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

Multivariate Distributions

Several random variables $(X_1, X_1, ..., X_n)$ can be associated with the same sample space. Then multivariate probability density/mass functions can be defined $f(x_1, x_2, .., x_n)$.

State Space for random vector (X_1, X_2)

$$A = \{(X_1(s), X_2(s)) | s \in S\} \subset \mathbb{R}^2$$

Joint Probability Density Function

Continuous random variables X_1, X_2 with probability measure $P : S \to [0, 1]$. $f : \mathbb{R}^2 \to [0, \infty)$ is the joint p.d.f if

$$\iint_A f(x_1, x_2) dx_1 dx_2 = 1 \quad \& \quad P_{(x_1, x_2)}(B) = \iint_B f(x_1, x_2) dx_1 dx_2$$

for $B \subset A$. The marginal p.d.f for X_1 is $f_1(x_1) = \int_{\mathbb{R}} f(x_1, x_2) dx_2$.

Multivariate Distributions

Joint Probability Mass Function

Discrete random variables X_1, X_2 with probability measure $P : S \to [0, 1]$.
$f : A \to [0, 1]$ is the joint p.m.f if

$$\sum_A f(x_1, x_2) = 1 \quad \& \quad P_{(x_1,x_2)}(B) = \sum_B f(x_1, x_2)$$

for $B \subset A$. The marginal p.m.f for X_1 is $f_1(x_1) = \sum_{x_2} f(x_1, x_2)$.

Covariance

The covariance of X_1 and X_2, two jointly distributed random variables is

$$cov(X_1, X_2) = E(X_1, X_2) - E(X_1)E(X_2).$$

If $cov(X_1, X_2) = 0$ the random variables are **uncorrelated**.

Generating Functions

Discrete random variable

X discrete random variable on state space $\{0, 1, 2, ...\}$ with p.m.f. $f(j) = \text{Prob}\{X = j\} = p_j$ where $\sum_{j=0}^{\infty} p_j = 1$.

$$\mu_X = E(X) = \sum_{j=0}^{\infty} j p_j \qquad \& \qquad \sigma_X^2 = E(X^2) - \mu_X^2 = \sum_{j=0}^{\infty} j^2 p_j - \mu_X^2.$$

Probability Generating Function

The p.g.f of X is the function

$$\mathcal{P}_X(t) = E(t^X) = \sum_{j=0}^{\infty} p_j t^j$$

for some $t \in \mathbb{R}$.

Since $\sum_{j=0}^{\infty} p_j = 1$ the sum converges absolutely if $|t| \leq 1$.

Generating Functions

Discrete random variable

Probability Generating Function

The p.g.f of X is the function

$$\mathcal{P}_X(t) = E(t^X) = \sum_{j=0}^{\infty} p_j t^j$$

for some $t \in \mathbb{R}$.

$\mathcal{P}(t)$ generates the probabilities associated with the distribution

$$\mathcal{P}_X(0) = p_0, \quad \mathcal{P}'_X(0) = p_1, \quad \mathcal{P}''_X(0) = 2!p_2, \quad \dots, \quad \mathcal{P}^{(k)}_X(0) = k!p_k$$

Mean and variance calculated from the p.g.f.

$$\mathcal{P}'_X(1) = \sum_{j=1}^{\infty} j p_j = E(X) = \mu_X \qquad \& \qquad \sigma^2_X = \mathcal{P}''_X(1) + \mathcal{P}'_X(1) - (\mathcal{P}'_X(1))^2$$

Generating Functions

Discrete random variable

Moment Generating Function

$$M_X(t) = E(e^{tx}) = \sum_{j=0}^{\infty} p_j e^{\,jt}$$

m.g.f. generates the moments $E(X^k)$ of the distribution of X.

$M_X(0) = 1, \quad M'_X(0) = E(X), \quad M''_X(0) = E(X^2), \ldots, \quad M_X^{(k)}(0) = E(X^k)$

Characteristic Function

$$\phi_X(t) = E(e^{itX}) = \sum_{j=0}^{\infty} p_j e^{ijt}$$

Cumulant Generating Function

$$K_X(t) = \ln[M_X(t)]$$

Generating Functions

Continuous random variable X with p.d.f f

Probability Generating Function

$$\mathcal{P}_X(t) = E(t^X) = \int_{\mathbb{R}} f(x) t^x dx$$

Moment Generating Function

$$M_X(t) = E(e^{tx}) = \int_{\mathbb{R}} f(x) e^{tx} dx$$

Characteristic Function

$$\phi_X(t) = E(e^{itX}) = \int_{\mathbb{R}} f(x) e^{ixt} dx$$

Cumulant Generating Function

$$K_X(t) = \ln[M_X(t)]$$

Generating Functions

Continuous random variable X with p.d.f f

Mean and variance calculated from the p.g.f., m.g.f, and c.g.f.

$$\mu_X = \mathcal{P}'_X(1) = M'_X(0) = K'_X(0)$$

$$\sigma_X^2 = \begin{cases} \mathcal{P}''_X(1) + \mathcal{P}'_X(1) - [\mathcal{P}'_X(1)]^2 \\ M''_X(0) - [M'_X(0)]^2 \\ K''_X(0) \end{cases}$$

Central Limit Theorem

Let $X_1, X_2, ..., X_n, ...,$ be a sequence of independent and identically distributed variables (*i.i.d*: n independent variables with the same distribution) with finite mean $|\mu| < \infty$ and standard deviation $0 < \sigma < \infty$.

$$W_n = \frac{\sum_{i=1}^{n} X_i/n - \mu}{\sigma/\sqrt{n}}$$

has a standard normal distribution for $n \to \infty$.

- Relates the sum of independent random variables to the normal dist.
- Even if the original variables themselves are not normally distributed
    - For any distribution of the X_n (as long as mean/variance finite), these amplified differences will be $\sim$ normally distributed for n large.
    - If the distribution is skewed and discrete - the sample size may need to be large for a good approximation
- This observation presumably explains why so many quantities are approximately normally distributed
    - The microscopic details are lost to the limit when we consider a macroscopic system in which the components act additively.