12 Lecture 4: Feb 2

Last time

- Linear algebra: vector and vector space, rank of a matrix
- Column space and Nullspace (JM Appendix A)

Today

- Probability review

Reference:

- Statistical Inference, 2nd Edition, by George Casella & Roger L. Berger
- Review of Probability Theory by Arian Maleki and Tom Do

Probability theory review

A few basic elements to define a probability on a set:

- **Sample space** S is the set that contains all possible outcomes of a particular experiment.
- An **event** is any collection of possible outcomes of an experiment, that is , any subset of S (including S itself).
- Event operations
  1. Union: The union of A and B, written $A \cup B$, is the set of elements that belong to either A or B or both:
  $$A \cup B = \{x : x \in A \text{ or } x \in B\}$$
  2. Intersection: The intersection of A and B, written $A \cap B$, is the set of elements that belong to both A and B:
  $$A \cap B = \{x : x \in A \text{ and } x \in B\}$$
  3. Complementation: The complement of A, written as A^c, is the set of all elements that are not in A:
  $$A^c = \{x : x \notin A\}.$$
- **Sigma algebra (or Borel field)**: A collection of subsets of S is called a sigma algebra (or Borel field), denoted by $\mathcal{B}$, if it satisfies the following three properties:
  1. $\varnothing \in \mathcal{B}$ (the empty set is an element of $\mathcal{B}$)
  2. If $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$ ($\mathcal{B}$ is closed under complementation).

3. If $A_1, A_2, \dots \in \mathcal{B}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$ ($\mathcal{B}$ is closed under countable unions).

- **Axioms of probability:** Given a sample space S and an associated sigma algebra $\mathcal{B}$, a *probability function* is a function $\Pr()$ with domain $\mathcal{B}$ that satisfies
  1. $\Pr(A) \geqslant 0$ for all $A \in \mathcal{B}$
  2. $\Pr(S) = 1$.
  3. If $A_1, A_2, \dots \in \mathcal{B}$ are pairwise disjoint, then $\Pr(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \Pr(A_i)$.

Properties:

If $\Pr()$ is a *probability function* and A and B are any sets in $\mathcal{B}$, then

- $\Pr(\varnothing) = 0$, where $\varnothing$ is the empty set
  Proof:
- $\Pr(A) \leqslant 1$
  Proof:
- $\Pr(A^c) = 1 - \Pr(A)$
  Proof:
- $\Pr(B \cap A^c) = \Pr(B) - \Pr(A \cap B)$
  Proof:
- $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
  Proof:
- $Pr(A \cup B) = \Pr(A) + \Pr(B \cap A^c) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
- If $A \subset B$, then $\Pr(A) \leqslant \Pr(B)$.
  Proof:

Conditional probability

Definition: If A and B are events in S, and $\Pr(B) > 0$, then the conditional probability of A given B, written $\Pr(A|B)$, is

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

Note that what happens in the conditional probability calculation is that B becomes the sample space: $\Pr(B|B) = 1$, in other words, $\Pr(A|B)$ is the probability measure of the event A after observing the occurrence of event B.

Definition: Two events A and B are statistically independent if $\Pr(A \cap B) = \Pr(A)\Pr(B)$. When A and B are independent events, then $\Pr(A|B) = \Pr(A)$ and the following pairs are also independent

- A and B^c
  proof:

- A^c and B
- A^c and B^c

Random variables

Definition: A random variable is a function from a sample space S into the real numbers.

Experiment	Random variable
Toss two dice	X = sum of the numbers
Toss a coin 25 times	X = number of heads in 25 tosses
Apply different amounts of fertilizer to corn plants	$X = yield/acre$

Suppose we have a sample space

$$S = \{s_1, \ldots, s_n\}$$

with a probability function Pr and we define a random variable X with range $\mathcal{X} = \{x_1, \ldots, x_m\}$. We can define a probability function $\Pr_X$ on $\mathcal{X}$ in the following way. We will observe $X = x_i$ if and only if the outcome of the random experiment is an $s_j \in S$ such that $X(s_j) = x_i$. Thus,

$$\Pr_X(X = x_i) = \Pr(\{s_j \in S : X(s_j) = x_j\}).$$

We will simply write $\Pr(X = x_i)$ rather than $\Pr_X(X = x_i)$.
A note on notation: Randon variables are often denoted with uppercase letters and the realized values of the variables (or its range) are denoted by corresponding lowercase letters.

Distribution functions

Definition: The cumulative distribution function or *cdf* of a random variable (r.v.) X, denoted by $F_X(x)$ is defined by

$$F_X(x) = \Pr(X \leqslant x), \text{ for all } x.$$

The function $F(x)$ is a cdf if and only if the following three conditions hold:

1. $\lim_{x\to\infty} F(x) = 1$.
2. $F(x)$ is a nondecreasing function of x.
3. $F(x)$ is right-continuous; that is, for every number x_0, $\lim_{x\downarrow x_0} = F(x_0)$.

Definition: A random variable X is continuous if $F(x)$ is a continuous function of x. A random variable X is discrete if $F(x)$ is a step function of x.

The following two statements are equivalent:

1. The random variables X and Y are identically distributed.
2. $F_X(x) = F_Y(x)$ for every x.

Density and mass functions

Definition: The probability mass function (pmf) of a discrete random variable X is given by

$$f_X(x) = \Pr(X = x) \text{ for all } x.$$

Example (Geometric probabilities) For the geometric distribution, we have the pmf

$$f_X(x) = \Pr(X = x) = \begin{cases} p(1-p)^{x-1} & \text{for } \mathrm{x} = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases}$$

Definition: The probability density function or *pdf*, $f_X(x)$, of a continuous random variable X is the function that satisfies

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt \quad \text{for all } x.$$

A note on notation: The expression "X has a distribution given by $F_X(x)$" is abbreviated symbolically by "$X \sim F_X(x)$", where we read the symbol "$\sim$" as " is distributed as".

Example (Logistic distribution) For the logistic distribution, we have

$$F_X(x) = \frac{1}{1 + e^{-x}}$$

and, hence,

$$f_X(x) = \frac{d}{dx} F_X(x) = \frac{e^{-x}}{(1 + e^{-x})^2}.$$

A function $f_X(x)$ is a pdf (or pmf) of a random variable X if and only if

1. $f_X(x) \geqslant 0$ for all x
2. $\sum_x f_X(x) = 1$ *(pmf)* *or* $\int_{-\infty}^{\infty} f_X(x)dx = 1$ *(pdf)*.

Expectations

The expected value, or expectation, of a random variable is merely its average value, where we speak of "average" value as one that is weighted according to the probability distribution.

Definition: The expected value or mean of a random variable $g(X)$, denoted by $\mathbf{E}(g(X))$, is

$$\mathbf{E}(g(X)) = \begin{cases} \int_{-\infty}^{\infty} g(x) f_X(x) dx & \text{if } X \text{ is continuous} \\ \sum_{x \in \mathcal{X}} g(x) f_X(x) = \sum_{x \in \mathcal{X}} g(x) \Pr(X = x) & \text{if } X \text{ is discrete,} \end{cases}$$

Exponential mean

Suppose $X \sim Exp(\lambda)$ distribution, that is, it has pdf given by

$$f_X(x) = \frac{1}{\lambda} e^{-x/\lambda}, \quad 0 \leqslant x < \infty, \quad \lambda > 0$$

Then $\mathbf{E}(X)$ is:

Binomial mean

IF X has binomial distribution, i.e. $X \sim binomial(n,p)$, its pmf is given by

$$\Pr(X=x) = \left(\begin{array}{c} \text{n} \\ \text{x} \end{array} \right) p^x(1-p)^{n-x}, \quad x = 0, 1, \ldots, n,$$

where n is a positive integer, $0 \leqslant p \leqslant 1$, and for every fixed pair n and p the pmf sums to 1. The expected value of a binomial random variable is then given by

$$\mathbf{E}\,(X) = \sum_{x=0}^{n} x \left(\begin{array}{c} \text{n} \\ \text{x} \end{array} \right) p^x(1-p)^{n-x}$$

Now, use the identity $x \left(\begin{array}{c} \text{n} \\ \text{x} \end{array} \right) = n \left(\begin{array}{c} \text{n - 1} \\ \text{x - 1} \end{array} \right)$ to derive the Expected value.

properties:

Let X be a random variable and let a, b and c be constants. Then for any functions $g_1(x)$ and $g_2(x)$ whose expectations exist,

1. $\mathbf{E}\,(a \cdot g_1(X) + b \cdot g_2(X) + c) = a\mathbf{E}\,(g_1(X)) + b\mathbf{E}\,(g_2(X)) + c.$
2. If $g_1(x) \geqslant 0$ for all x, then $\mathbf{E}\,(g_1(X)) \geqslant 0$.
3. If $g_1(x) \geqslant g_2(x)$ for all x, then $\mathbf{E}\,(g_1(X)) \geqslant \mathbf{E}\,(g_2(X))$.
4. If $a \leqslant g_1(x) \leqslant b$ for all x, then $a \leqslant \mathbf{E}\,(g_1(X)) \leqslant b$.

Moments

The various moments of a distribution are an important class of expectations.

Definition: For each integer n, the n^{th} moment of X (or $F_X(x)$), μ_n', is

$$\mu_n' = \mathbf{E}\,(X^n).$$

The n^{th} central moment of X, μ_n, is

$$\mu_n = \mathbf{E}\,((X-\mu)^n),$$

where $\mu = \mu_1' = \mathbf{E}\,(X)$.

Variance

Definition: The variance of a random variable X is its second central moment, $\mathbf{Var}\,(X) = \mathbf{E}\,((X - EX)^2)$. The positive square root of $\mathbf{Var}\,(X)$ is the standard deviation of X.

Exponential variance

Let X have the exponential(λ) distribution, $X \sim Exp(\lambda)$. Then the variance of X is properties

1. $\mathbf{Var}(aX+b) = a^2\mathbf{Var}(X)$.
   proof:

2. $\mathbf{Var}(X) = \mathbf{E}(X^2) - (\mathbf{E}(X))^2$.
   proof:

Moment generating function

Definition: Let X be a random variable with cdf F_X. The moment generating function or mgf of X (or F_X), denoted by $M_X(t)$, is

$$M_X(t) = \mathbf{E}\left(e^{tX}\right),$$

provided that the expectation exists for t in some neighborhood of 0. That is, there exists an $h > 0$ such that for all t in $-h < t < h$, $\mathbf{E}\left(e^{tX}\right)$ exists. If the expectation does not exist in a neighborhood of 0, we say that the moment generating function does not exist.

Property: If X has mgf $M_X(t)$, then

$$\mathbf{E}(X^n) = M_X^{(n)}(0),$$

where we define

$$M_X^{(n)}(0) = \left.\frac{d^n}{dt^n}M_X(t)\right|_{t=0}.$$

Some common random variables

Discrete random variables

- $X \sim \mathit{Bernoulli}(p)$ (where $0 \leqslant p \leqslant 1$):

$$\Pr(x) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \end{cases}$$

- $X \sim \mathit{Binomial}(n,p)$ (where $0 \leqslant p \leqslant 1$):

$$\Pr(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

- $X \sim \mathit{Geometric}(p)$ (where $0 \leqslant p \leqslant 1$):

$$\Pr(x) = p(1-p)^{x-1}$$

- $X \sim \mathit{Poisson}(\lambda)$ (where $\lambda > 0$):

$$\Pr(x) = e^{-\lambda}\frac{\lambda^x}{x!}$$

Continuous random variables

- $X \sim Uniform(a, b)$ (where $a < b$):

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

- $X \sim Exponential(\lambda)$ (where $\lambda > 0$):

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } \ x \geqslant 0 \\ 0 & \text{otherwise} \end{cases}$$

- $X \sim Normal(\mu, \sigma^2)$:

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

The following table provides a summary of some of the properties of these distributions.

Distribution	PDF or PMF	Mean	Variance
$Bernoulli(p)$	$\begin{cases} p & \text{if } \ x = 1 \\ 1-p & \text{if } \ x = 0 \end{cases}$	p	$p(1-p)$
$Binomial(n, p)$	$\begin{pmatrix} n \\ x \end{pmatrix} p^x(1-p)^{n-x}$, for $0 \leqslant k \leqslant n$	np	$np(1-p)$
$Geometric(p)$	$p(1-p)^{x-1}$, for $k = 1, 2, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$Poisson(\lambda)$	$e^{-\lambda}\frac{\lambda^x}{x!}$, for $k = 1, 2, \ldots$	λ	λ
$Uniform(a, b)$	$\frac{1}{b-a}I(a \leqslant x \leqslant b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$Gaussian(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$	μ	σ^2
$Exponential(\lambda)$	$\lambda e^{-\lambda x}I(x \geqslant 0)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$