MATH 323 Fall 2018
Course Summary

1. The basics of probability.
   (i) Review of set theory notation.
      - intersection, union, complement and how they combine
      - manipulating multiple events
      - partitions

   (ii) Sample spaces and events.
      - definitions and terminology

(iii) The probability axioms and their consequences.

- definitions
- axioms
- corollaries
- general addition rule
- probability tables

(iv) Probability spaces with equally likely outcomes.

- Decomposition of S into equally likely sample outcomes
- Calculations of the form

$$P(A) = \frac{n_A}{n_S}$$

(v) Combinatorial probability.

- multiplication principle
- selecting with and without replacement
- permutations
- multinomial coefficients and partitioning sets
- combinations
- binary sequences
- hypergeometric selection

(vi) Conditional probability and independence.

- concept of conditional probability
- definition
- properties
- independence
- mutual independence
- general multiplication (or chain) rule

(vii) The Theorem of Total Probability.

- ‘proof’ by partitioning
- consequences
- probability trees

(viii) Bayes Theorem.

- ‘proof’ by definition of conditional probability
- interpretation and consequences
- probability trees

2. Random variables and probability distributions.

   (i) Random variables.

   - definition
   - elementary examples

   (ii) Discrete random variables and distributions

   - pmfs $p(y) = P(Y = y)$
   - basic properties

$$0 \leq p(y) \leq 1 \qquad \sum_y p(y) = 1$$

   - cdfs $F(y) = P(Y \leq y)$
   - basic properties: non-decreasing, right-continuous step function

$$F(-\infty) = 0 \qquad F(\infty) = 1$$

   - basic computations

(iii) Continuous random variables and distributions:

- cdfs $F(y) = P(Y \leq y)$
- basic properties: increasing, continuous

$$F(-\infty) = 0 \quad F(\infty) = 1$$

- pdfs $f(y) = \frac{dF(y)}{dy}$
- basic properties:

$$f(y) \geq 0 \qquad \int_{-\infty}^{\infty} f(y)\, dy = 1$$

- special cases: pdfs defined in a piecewise fashion

(iv) Moments:

- general expectations

$$\mathbb{E}[g(Y)] = \begin{cases} \sum_y g(y)p(y) & \text{discrete} \\ \\ \int_{-\infty}^{\infty} g(y)f(y)\,dy & \text{continuous} \end{cases}$$

- expectation and variance.
- basic properties
- linear transformations

(v) Moment generating functions (mgfs):

- derivation and uses.

$$m(t) = \mathbb{E}[e^{tY}] = \begin{cases} \sum_{y} e^{ty} p(y) & \text{discrete} \\ \\ \int_{-\infty}^{\infty} e^{ty} f(y)\, dy & \text{continuous} \end{cases}$$

$$t \in (-b, b)$$

$$m(0) = 1$$

$$m^{(r)}(0) = \frac{d^r}{dt^r} \{m(t)\}_{t=0} = \mathbb{E}[Y^r]$$

and uniqueness.

- other related generating functions.
- pgf (discrete case)

$$G(t) = \mathbb{E}[t^Y] = \sum_y t^y p(y) \qquad t \in (1-b, 1+b)$$

$$G(1) = 1$$

$$G^{(r)}(0) = \frac{d^r}{dt^r}\{G(t)\}_{t=0} = r!p(r)$$

$$G^{(r)}(1) = \frac{d^r}{dt^r}\{G(t)\}_{t=1} = \mathbb{E}[Y(Y-1)(Y-2)\ldots(Y-r+1)]$$

- pgf, fmgf (continuous case)

$$G(t) = \mathbb{E}[t^Y] = \int_{-\infty}^{\infty} t^y f(y)\, dy \qquad t \in (1-b, 1+b)$$

$$G(1) = 1$$

$$G^{(r)}(1) = \frac{d^r}{dt^r}\left\{G(t)\right\}_{t=1} = \mathbb{E}[Y(Y-1)(Y-2)\dots(Y-r+1)]$$

$$G(t) = m(\ln t)$$

(vi) Named distributions:

- discrete uniform,
- hypergeometric,
- binomial,
- geometric,
- negative binomial,
- Poisson,
- continuous uniform,
- gamma,
- exponential,
- chi-squared,
- beta,
- Normal.

Should know (where possible) the experimental context, connections between distributions.

3. Probability calculation methods.

- Transformations in one dimension: $U = h(Y)$
  - Discrete case:

$$p_U(u) = P(U = u) = \sum_{y \in A_u} p_Y(y)$$

where

$$A_u\{y : h(y) = u\}$$

that is: sum the probabilities over all y points that map onto the value u.

If h is 1-1: straightforward.

- Continuous case: first principles – start with $F_U(u)$

$$F_U(u) = P(U \leq u) = P(h(Y) \leq u) = \int_{A_u} f_Y(y)\, dy$$

where

$$A_u = \{y : h(y) \leq u\}.$$

If h is monotonic, h^{-1} is well defined, so write

$$P(h(U) \leq u) = \begin{cases} P(Y \leq h^{-1}(u)) & h \text{ increasing} \\ \\ P(Y \geq h^{-1}(u)) & h \text{ decreasing} \end{cases}$$

Could also use the derived "Jacobian" result for the pdf:

$$f_U(u) = f_Y(h^{-1}(u)) \left| \frac{dh^{-1}(u)}{du} \right|$$

to cover both cases. The term

$$\left| \frac{dh^{-1}(u)}{du} \right|$$

is the *Jacobian* of the transformation.

- Techniques for sums of random variables.

$$Y = Y_1 + Y_2$$

with Y_1 and Y_2 independent.

Direct calculations using Convolution results.

- discrete case is simply the Theorem of Total Probability

$$p_Y(y) = P(Y = y) = \sum\sum_{A_y} p_{Y_1}(y_1)p_{Y_2}(y_2)$$

where

$$A_y = \{(y_1, y_2) : y_1 + y_2 = y\}$$

but also

$$\begin{aligned} p_Y(y) = P(Y = y) &= \sum_{y_1=-\infty}^{\infty} P(Y = y|Y_1 = y_1)P(Y_1 = y_1) \\ &= \sum_{y_1=-\infty}^{\infty} P(Y_2 = y - y_1)P(Y_1 = y_1) \\ &= \sum_{y_1=-\infty}^{\infty} p_{Y_1}(y_1)p_{Y_2}(y - y_1) \end{aligned}$$

- Continuous case is the analogue, but we start formally with $F_Y(y)$

$$F_Y(y) = P(Y \leq y) = \iint_{A_y} f_{Y_1}(y_1) f_{Y_2}(y_2)\, dy_2 dy_1$$

where

$$A_y = \{(y_1, y_2) : y_1 + y_2 \leq y\}$$

that is

$$F_Y(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{y-y_1} f_{Y_1}(y_1) f_{Y_2}(y_2)\, dy_2 dy_1$$

Differentiating wrt y gives

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y_1}(y_1) f_{Y_2}(y - y_1)\, dy_1$$

4. Multivariate distributions.

- Joint distributions: discrete case joint pmf, joint cdf

$$p_{Y_1,Y_2}(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2)$$

  and

$$F_{Y_1,Y_2}(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2) = \sum_{t_1=-\infty}^{y_1} \sum_{t_2=-\infty}^{y_2} p_{Y_1,Y_2}(t_1, t_2)$$

  Properties of both.

- Joint distributions: continuous case joint cdf, joint pdf

$$F_{Y_1,Y_2}(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2)$$

for any pair of real numbers (y_1, y_2).

  - "starts at zero"

$$\lim_{y_1 \longrightarrow -\infty} \lim_{y_2 \longrightarrow -\infty} F_{Y_1,Y_2}(y_1, y_2) = 0$$

and

$$\lim_{y_1 \longrightarrow -\infty} F_{Y_1,Y_2}(y_1, y_2) = 0 \qquad \forall y_2$$

and

$$\lim_{y_2 \longrightarrow -\infty} F_{Y_1,Y_2}(y_1, y_2) = 0 \qquad \forall y_1.$$

- "ends at one"

$$\lim_{y_1 \longrightarrow \infty} \lim_{y_2 \longrightarrow \infty} F_{Y_1,Y_2}(y_1, y_2) = 1$$

- "non-decreasing in y_1 and y_2 in between"

$$F_{Y_1,Y_2}(y_1, y_2) \leq F_{Y_1,Y_2}(y_1 + c, y_2)$$

$$F_{Y_1,Y_2}(y_1, y_2) \leq F_{Y_1,Y_2}(y_1, y_2 + c)$$

for all y_1, y_2, and any $c > 0$.

Relationship:

$$F_{Y_1,Y_2}(y_1, y_2) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f_{Y_1,Y_2}(t_1, t_2)\, dt_2 dt_1$$

and

$$f_{Y_1,Y_2}(y_1, y_2) = \frac{\partial^2}{\partial y_1 \partial y_2} \{F_{Y_1,Y_2}(y_1, y_2)\}$$

Probability calculations of the sort

$$P(g(Y_1, Y_2) \in A) = \iint_A f_{Y_1}(y_1) f_{Y_2}(y_2)\, dy_2 dy_1$$

for some transformation $g(.,.)$.

Key is to identify (eg sketch) the region of integration.

- Marginal cdfs

$$F_{Y_1}(y_1) = F_{Y_1,Y_2}(y_1,\infty) \qquad F_{Y_2}(y_2) = F_{Y_1,Y_2}(\infty,y_2)$$

and pmfs

$$p_{Y_1}(y_1) = \sum_{y_2=-\infty}^{\infty} p_{Y_1,Y_2}(y_1,y_2)$$

and pdfs

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1,y_2)\,dy_2$$

Marginal distributions are regular distributions, so can compute all the standard quantities (eg expectations, generating functions etc).

- Conditional cdfs

$$F_{Y_1|Y_2}(y_1|y_2) = P(Y_1 \leq y_1|Y_2 = y_2)$$

and pmfs

$$p_{Y_1|Y_2}(y_1|y_2) = \frac{p_{Y_1,Y_2}(y_1, y_2)}{p_{Y_2}(y_2)}$$

and pdfs.

$$f_{Y_1|Y_2}(y_1|y_2) = \frac{f_{Y_1,Y_2}(y_1, y_2)}{f_{Y_2}(y_2)}$$

Conditional distributions are regular distributions, so can compute all the standard quantities (eg expectations, generating functions etc).

- Independence of random variables.

$$F_{Y_1,Y_2}(y_1,y_2) = F_{Y_1}(y_1)F_{Y_2}(y_1)$$

or

$$p_{Y_1,Y_2}(y_1,y_2) = p_{Y_1}(y_1)p_{Y_2}(y_1)$$

or

$$f_{Y_1,Y_2}(y_1,y_2) = f_{Y_1}(y_1)f_{Y_2}(y_1)$$

for all $(y_1, y_2) \in \mathbb{R}^2$.

Quick check to exclude possibility of independence: is the support of the joint pmf/pdf a Cartesian product? If not, variables are not independent.

- Linear combinations of random variables

$$U_1 = \sum_{i=1}^{n} a_i Y_i \qquad U_2 = \sum_{j=1}^{m} b_j X_j$$

for real constants $a_1, \ldots, a_n$ and $b_1, \ldots, b_m$.

$$\text{Expectations: } \mathbb{E}[U_1] = \sum_{i=1}^{n} a_i \mathbb{E}[Y_i] = \sum_{i=1}^{n} a_i \mu_i$$

$$\text{Variances: } \mathbb{V}[U_1] = \sum_{i=1}^{n} a_i^2 \mathbb{V}[Y_i] + 2 \sum_{j=2}^{n} \sum_{i=1}^{j-1} a_i a_j \mathrm{Cov}[Y_i, Y_j]$$

$$\text{Covariances: } \mathrm{Cov}[U_1, U_2] = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \mathrm{Cov}[Y_i, X_j]$$

- mgfs for sums of independent random variables
  If $Y_1, \ldots, Y_n$ are independent, then

$$Y = \sum_{i=1}^{n} Y_i$$

then

$$m_Y(t) = \prod_{i=1}^{n} m_{Y_i}(t).$$

- Covariance and correlation: particular multivariate expectations.

$$\mathrm{Cov}[Y_1, Y_2] = \mathbb{E}[(Y_1 - \mu_1)(Y_2 - \mu_2)] = \mathbb{E}[Y_1 Y_2] - \mu_1 \mu_2$$

$$\mathrm{Corr}[Y_1, Y_2] = \frac{\mathrm{Cov}[Y_1, Y_2]}{\sqrt{\mathbb{V}[Y_1] V[Y_2]}}$$

These are measures of dependence or association.

5. Probability inequalities and theorems.
   - Markov's inequality; Chebychev's inequality.
   - The Weak Law of Large Numbers.
     - Sample mean converges to theoretical expectation as $n \longrightarrow \infty$.

$$\overline{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

then

$$\overline{Y}_n \xrightarrow{p} \mu$$

as $n \longrightarrow \infty$.

- The Central Limit Theorem and approximation methods.
  - Sums of independent and identically distributed random variables are approximately Normally distributed as $n \longrightarrow \infty$.

The quantity

$$U_n = \frac{\sqrt{n}(\overline{Y}_n - \mu)}{\sigma}$$

converges in distribution to a *standard normal distribution*.

Alternately

$$U_n = \frac{(\overline{Y}_n - \mu)}{\sigma/\sqrt{n}}$$

or, multiplying numerator and denominator by n

$$U_n = \frac{\left(\sum_{i=1}^{n} Y_i - n\mu\right)}{\sqrt{n}\sigma}$$

- $\overline{Y}_n$ is approximately distributed as

$$Normal(\mu, \sigma^2/n)$$

- If

$$S_n = \sum_{i=1}^{n} Y_i$$

then S_n is approximately distributed as

$$Normal(n\mu, n\sigma^2)$$

The Core

Core methods

1. Axioms & Probability Calculations
2. Conditional Probability
3. Theorem of Total Probability and Bayes Theorem
4. pmfs, cdfs, pdfs: properties and calculations
5. Expectations
6. Generating Functions: definitions, properties and uses
7. Standard distributions and their connections
8. Transformations
9. Multivariate distributions: properties and calculations (including convolutions, marginalization, expectations)
10. Normal approximations

Exam

Instructions

- Exam will last three hours.
- Exam will contain 5 questions, 20 marks each.
- Rescaling of the final mark may occur.
    - my aim is get the average between 75% and 80%.
- Answer the questions in the booklet provided.

Instructions (cont.)

- Show your working if asked to do so.
- You may quote without proof results from the formula sheet.
- You may leave combinatorial results in terms of binomial coefficients unless asked to compute numerically.

Advice

- Try to figure out the most efficient way to obtain the solution.
- You will not need to do very long calculations IF you figure out the correct approach.
- If in doubt, go back to first principles.
- If you feel a question is ambiguous, note the claimed ambiguity in your solution and then proceed to answer according to your interpretation of what the question is asking.