Probability IA Lent

1 Basic Probability

- Probability space: Axioms
  - Sample space (Ω) consists of outcomes ω
  - σ-algebra ($\mathcal{F}$) consists of events:
    $\Omega \in \mathcal{F}$
    Closed under complement and countable union
  - Probability measure ($\mathbb{P} : \mathcal{F} \to [0,1]$): $\mathbb{P}(\Omega) = 1$
    Countable additivity: disjoint sequence (A_n),

$$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$$

- Equally likely outcomes in finite sample space: probability given by counting:

$$P(A) = \frac{|A|}{|\Omega|}$$

- Combinatorial analysis:
  - Addition and multiplication rule
  - Permutation and combination
  - Multinomial coefficient
- Stirling's formula:

$$n! \sim \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}$$

- Countable subadditivity:

$$P(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n)$$

- Continuity of probability measures: for non-decreasing sequence of events $A_1 \subseteq A_2 \subseteq A_3 \ldots$:

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} P(A_n)$$

- Inclusion-exclusion principle: useful for problems with symmetry, i.e. probability of an intersection of any n events is the same:

$$
\begin{aligned}
P\left(\bigcup_{j=1}^{n} A_j\right) &= \sum_{j=1}^{n}(-1)^{j+1} \sum_{1\leq i_1\leq \ldots i_j\leq n} P(A_{i_1}\cap\cdots\cap A_{i_j}) \\
&= \sum_{j=1}^{n}(-1)^{j+1}\binom{n}{j} P(A_1\cap\cdots\cap A_j)
\end{aligned}
$$

- Bonferroni Inequalities

2 Independence and Conditional Probability

- Independence:

$$P(A|B) = P(A) \text{ or}$$

The events $A_1, A_2, \ldots A_n$ are independent iff for any $k \geq 2$ and $1 \leq i_1 < i_2 < i_3 < \ldots i_k \leq n$

$$P(A_{i_1}\cap A_{i_2}\cap A_{i_3}\cap\cdots\cap A_{i_k}) = P(A_{i_1})P(A_{i_2})P(A_{i_3})\ldots P(A_{i_k})$$

- Conditional Probability:

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

- Law of Total Probability: Given a partition $(B_n)_{n\geq 1}$ of Ω,

$$P(A) = \sum_{n=1}^{\infty} P(A|B_n)P(B_n)$$

- Bayes' Theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

3 Discrete Probability Distributions

- Bernoulli: $\mathrm{Bern}(p) = \mathrm{Bin}(1,p)$

$$p_1 = p, p_0 = 1-p$$

- Binomial: Bin(n, p)

$$p_k = \binom{n}{k} p^k (1-p)^{n-k}, \text{ where } k \geq 0$$

- Poisson: Poisson$(\lambda), \lambda \in (0, \infty)$

$$p_k = \frac{e^{-\lambda}\lambda^k}{k!}, \text{ where } k \geq 0$$

- Geometric: Geom(p)

$$p_k = (1-p)^{k-1}p, \text{ where } k \geq 1$$

- Binomial approaches Poisson:

$$\text{Bin}(N, p) \to \text{Poisson}(Np)$$

4 Random Variable

- Random Variable: function that assigns (say real) values to each outcome

$$X : \Omega \to \Omega_X (= \mathbb{R} \text{ or } \mathbb{R}^n)$$

- Discrete random variable if Ω is countable (Random variable in general probability space need $\{\omega : X(\omega) < x\} \in \mathcal{F}$ for all x)
- Indicator function:

$$1_A(\omega) = \begin{cases} 1, \text{ if } \omega \in A \\ 0, \text{ if } \omega \notin A \end{cases}$$

- (Cumulative/probability) distribution function (for real-valued rv):

$$F_X(x) = P(X \leq x)$$

  - Piecewise constant
  - Non-decreasing
  - Right continuous
  - $\lim_{x \to -\infty} F_X(x) = 0$

  - $\lim_{x \to +\infty} F_X(x) = 1$

  (This is equivalent to specifying the weights (p_x) of the rv, i.e. specifying the probability mass function)

- Independence:

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

  Any n random variables $X_1, X_2, \dots X_n$ are independent iff

  $$P(X_1 = x_1, X_2 = x_2, \dots X_n = x_n) = P(X_1 = x_1)P(X_2 = x_2)\dots P(X_n = x_n)$$

  (This implies any subset of the rvs are independent)

- Functions of random variables: new random variable $g(X)$ defined by

$$g(X)(\omega) = g(X(\omega))$$

5 Expectation

- Expectation: Probability weighted sum of values

$$\begin{aligned}\mathrm{E}(X) &= \sum_{\omega \in \Omega} X(\omega)P(\omega) \\ &= \sum_{x \in \Omega_X} xP(X = x)\end{aligned}$$

  Defined iff the expectations the positive X and negative X are not both infinite ($\mathrm{E}(X_+)$, $\mathrm{E}(X_-)$)

- Integrable: rv X has $\mathrm{E}(|X|) \leq \infty$

- Properties of expectation: positive definite, linear, multiplicative if independent, $\mathrm{E}(1(A)) = P(A)$

- Moments: n-th moment $= \mathrm{E}(X^n)$

- Variance:

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}((X - \mathrm{E}(X))^2) \\ &= \mathrm{E}(X^2) - (\mathrm{E}(X))^2\end{aligned}$$

- Covariance:

$$\mathrm{Cov}(X,Y) = \mathrm{E}[(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))]$$

- Correlation:

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

6 Inequalities

- Markov's Inequality: Given non-negative rv X, and $\lambda > 0$

$$P(X \geq \lambda) \leq \frac{\mathrm{E}(X)}{\lambda}$$

- Chebyshev's Inequality: Given rv X with finite $\mathrm{E}(X)$, and $\lambda > 0$

$$P(|X - \mathrm{E}(X)| \geq \lambda) \leq \frac{\mathrm{Var}(X)}{\lambda^2}$$

- Cauchy-Schwarz:

$$\mathrm{E}(|XY|) \leq \sqrt{\mathrm{E}(X^2)\mathrm{E}(Y^2)}$$

- Jensen's Inequality: Given convex function f on I, X integrable on I

$$f(\mathrm{E}(X)) \leq \mathrm{E}(f(X))$$

  AM/GM then follows

7 Random walk (L9)

- Definition

$$X_n = x + Y_1 + Y_2 + \cdots + Y_n, \text{ where } Y_i \text{ are iid rvs}$$

- Gambler's ruin, Expected time of absorption ...

8 Conditional Expectation

- Definition
$$E(X|B) = \frac{E(X \cdot 1(B))}{P(B)}$$
- Law of Total Expectation: Given (Ω_i) a partition of Ω,
$$E(X) = \sum E(X|\Omega_i)P(\Omega_i)$$
- (Joint distribution, Marginal distribution, Conditional distribution)
- Convolution: given X, Y independent, $P(X + Y = z) = \sum_y P(X = z - y)P(Y = y)$
- Conditional expectation of X given Y (a random variable in terms of Y): if $E(X|Y = y) = g(y)$, then $E(X|Y) = g(Y)$
- $$E(X) = E_Y(E(X|Y))$$

9 Probability generating function (L9)

- Definition:
$$G_X(t) = \mathrm{E}(t^X) = \sum p_k t^k$$
- Can recover probability distribution by differentiation and put 0
- pgf of $S = \sum_{i=1}^n X_i$, where X_i independent, is
$$F_S(t) = \prod_i F_{X_i}(t)$$
- Random sum (N) of iid random variables (X_i): $\sum_{i=1}^N X_i$ is $F_N(F_{X_1}(t))$

10 Branching processes (L11)

- Galton-Watson Process
- Condition on first step, set number of branches
- Extinction probability

11 Continuous Random Variables

- X is a continuous rv if distribution function F_X is also left continuous
- Probability distribution function continuous $\implies$ X is continuous
- Probability distribution function differentiable $\implies$ X is absolutely continuous (then have probability density function)
- Probability density function (pdf) $f(x)$ satisfies

$$f(x) \geq 0 \text{ for all } x \in \mathbb{R}, \qquad \int_{\mathbb{R}} f(x)\mathrm{d}x = 1$$

- Examples of distributions (L13)
  - Uniform distribution: $X \sim U[a,b]$

$$f_X(x) = \frac{1}{b-a} 1_{[a,b]}(x)$$

  - Exponential, memoryless property: $X \sim Exp(\lambda)$

$$f_X(x) = \lambda e^{-\lambda x} 1_{\{x>0\}}(x)$$

  - Gamma distribution (generalised/sum of exponential) $X \sim \Gamma(\lambda, n)$, where $\lambda > 0, n \in \mathbb{N}$:

$$f_X(x) = e^{-\lambda x} \lambda^n \frac{x^{n-1}}{(n-1)!}, \text{ for } x \geqslant 0$$

  - Standard Cauchy distribution $X \sim \text{Cauchy}(0,1)$

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

  - Gaussian (standard normal) Normal: $X \sim N(\mu, \sigma^2)$

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

    Distribution function of standard normal $N(0,1)$: $\Phi(x)$

- Expectation ($E(X) = \int_{\mathbb{R}} x f_X(x)\mathrm{d}x$) and variance ($Var(X) = E[(X - E(X))^2]$) of continuous rv

12 Multivariate rv

- Joint distribution of rvs $X_1, \dots X_n$

$$F(x_1, \dots, x_n) = P(X_1 \leqslant x_1, \dots X_n \leqslant x_n)$$

- Joint density = mixed derivative
- Random variables are independent if and only if the density function factorises
- Marginal density = density of one rv (integrate everything else)
- Convolution (for independent rvs):

$$f_{X+Y}(z) = f_X * f_Y(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) \mathrm{d}y$$

- Conditional density:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

- Conditional expectation, law of total probability

13 Transformation of rv

- One-dimensional rv:

$$f_Y(y) = f_X(x) \left| \frac{\mathrm{d}x}{\mathrm{d}y} \right|$$

- Multi-dimensional rv:

$$f_Y(y) = f_X(x) \, |J|$$

where J is the Jacobian,

$$J = \det\left(\frac{\partial x_i}{\partial y_i}\right)_{i,j=1}^{d}$$

Order Statistics of a random sample (L15)

14 Simulation of random variables (L17)

- Want to generate rv X with distribution function $F(x)$:
  Using $U \sim U[0,1]$, then $F^{-1}(U)$ has same distribution as X

- Rejection Sampling:
  Given $A \subseteq [0,1]^d$ (d-dim unit cube), want to generate rv X in $\mathbb{R}^d$ with density function $f_X(x) = \frac{1_A(x)}{|A|}$ ('uniform' on A)

  Use $U[0,1]$, generate a sequence of d-dimensional rv (points in unit cube) $(U_n)_{n\in\mathbb{N}}$, where each point is

  $$U_n = (U_{1,n}, U_{2,n}, \ldots U_{d,n})$$

  Take $X = U_N$, where $N = \min$

- Bounded density ...

15 Moment Generating Function (L17)

- mgf of continuous rv X

  $$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_X(x) \mathrm{d}x$$

- Given $X_1, \ldots X_n$ independent,

  $$M_{X_1+X_2+\cdots+X_n}(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

- Uniquely determine distribution if mgf is finite in some neighbourhood of the origin

- Continuity of mgf:
  Suppose X is a rv, and (X_n) a sequence of cts rv, if $M_{X_n}(\theta) \to M_X(\theta) \forall\theta \in \mathbb{R}$ and $M_X(\theta)$ is finite for some $\theta \neq 0$, then X_n converges to X in distribution

- Multivariate mgf:
  Suppose $X = (X_1, \ldots, X_n)$ is a rv in $\mathbb{R}^n$, then mgf of X is

  $$E(e^{\theta^T X})$$

16 Gaussian: Multivariate

- X in $\mathbb{R}$ is Gaussian if

$$X = \mu + \sigma Z, \text{ where } Z \sim N(0,1)$$

- $X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$ a rv in $\mathbb{R}^n$ is Gaussian

  if $u^T X = \sum_{i=1}^n u_i X_i$ is Gaussian for any $u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \in \mathbb{R}^n$

- Mean and variance

$$\mu = E(X) = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_n) \end{pmatrix}$$

$$\begin{aligned} V &= \text{Var(X)} \\ &= E((X-\mu)(X-\mu)^T) \\ &= E \begin{pmatrix} (X_1-\mu_1)^2 & (X_1-\mu_1)(X_2-\mu_2) & \cdots & (X_1-\mu_1)(X_n-\mu_n) \\ \vdots & (X_2-\mu_2)^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \cdots & \cdots & (X_n-\mu_n)^2 \end{pmatrix} \\ &= (\text{Cov}(X_i, X_j)) \end{aligned}$$

- mgf of X is

$$M_X(\lambda) = E(e^{\lambda^T X}) = e^{\lambda^T \mu + \frac{\lambda^T V \lambda}{2}}$$

  (uniquely determined by mean and variance), written $X \sim N(\mu, V)$

- Can use $Z = (Z_1, \ldots Z_n)$, where (Z_i) iid $\sim N(0,1)$ in $\mathbb{R}^n$ to construct any n-dim Gaussian vector: $X = \mu + \sigma Z \sim N(\mu, V)$ (σ is the square root of matrix V)

- Density of X:
$X = \mu + \sigma Z$, if V is positive definite (invertible), have density function

$$\begin{aligned} f_X(x) &= f_Z(z)|J| \\ &= \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}} e^{-z_i^2/2} \right) |\det \sigma^{-1}| \\ &= \frac{1}{(2\pi)^{n/2}} e^{-z^T z/2} \frac{1}{(\det V)^{1/2}} \\ &= \frac{1}{(2\pi)^{n/2} (\det V)^{1/2}} e^{\frac{-(x-\mu)^T V (x-\mu)}{2}} \end{aligned}$$

In general: non-negative definite (can have zero eigenvalues): diagonalise V to get

$$V = \left(\begin{array}{c|c} U & 0 \\ \hline 0 & 0 \end{array} \right),$$

where U is an $m \times m$ positive definite matrix. Write $\mu = \begin{pmatrix} \lambda \\ \nu \end{pmatrix}$, where $\lambda \in \mathbb{R}^m$. Then $X = \begin{pmatrix} Y \\ \nu \end{pmatrix}$, and density of Y is given by

$$f_Y(y) = \frac{1}{(2\pi)^{m/2} (\det U)^{1/2}} e^{\frac{-(y-\lambda)^T U (y-\lambda)}{2}}$$

- Bivariate normal
- (Multivariate Central Limit Theorem)

17 Limit Theorems

- Weak law of large numbers:
Given $S_n = X_1 + \cdots + X_n$, where (X_i) iid with $E(X_1) = \mu < \infty$, then for all $\varepsilon > 0$, as $n \to \infty$

$$P\left(\left| \frac{S_n}{n} - \mu \right| > \varepsilon \right) \to 0$$

(converge in probability)

- Strong law of large numbers:
Given $S_n = X_1 + \cdots + X_n$, where (X_i) integrable iid with $E(X_1) = \mu < \infty$, then

$$P\left(\lim_{n \to \infty} \frac{S_n}{n} = \mu \right) = 1$$

(almost surely/converge with prob 1)

- Central Limit Theorem:
  Given $S_n = X_1 + \cdots + X_n$, where (X_i) iid with $E(X_1) = \mu < \infty$ and $\text{Var}(X_1) = \sigma^2 < \infty$, then

$$P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) \to \Phi(x)$$

$$\text{i.e. } S_n \sim N(n\mu, n\sigma^2)$$

18 Geometric Probs

- Box Muller Transform
- Buffon's needle
- Bertrand's paradox: defn of random