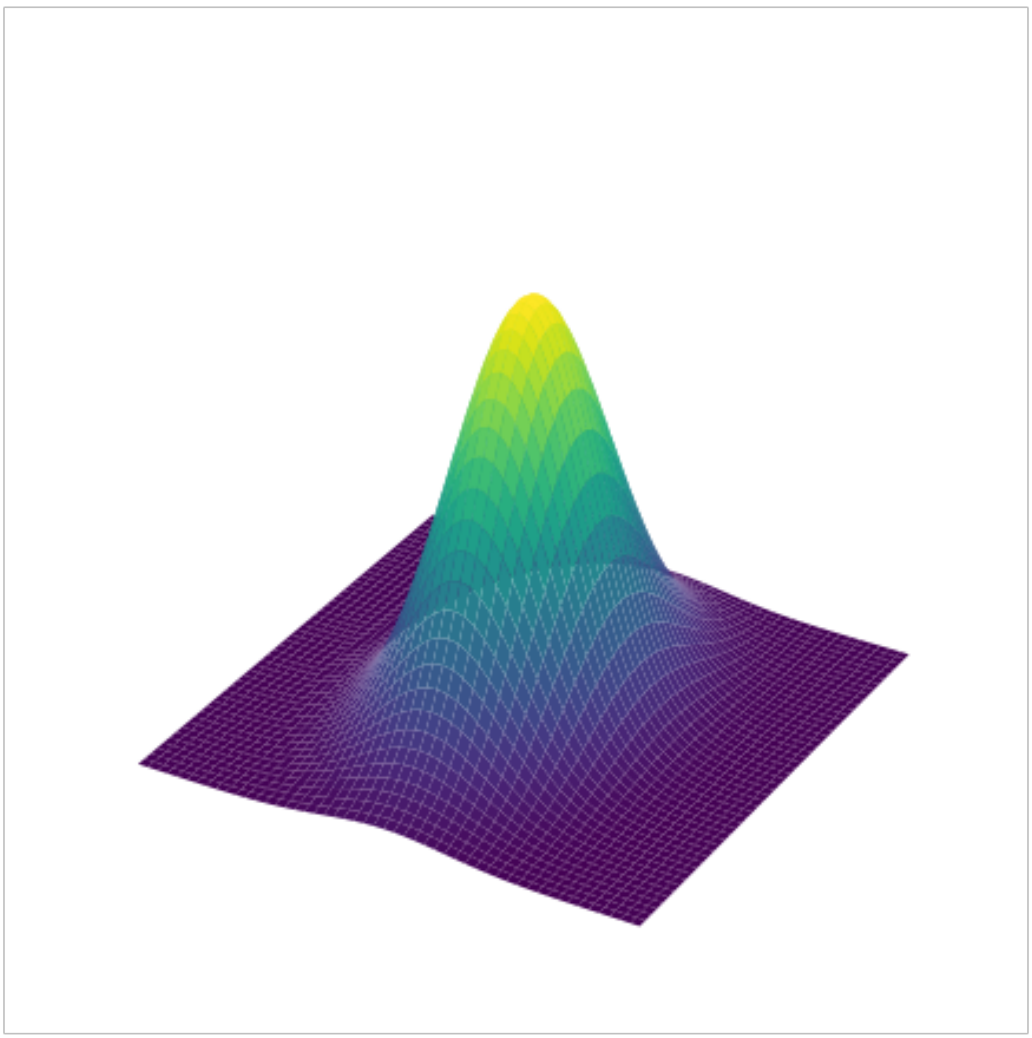
The University of Melbourne

Semester 2 2020

Course Notes

MAST20004 Probability



MAST20004 Probability

Probability Axioms	$\mathbb{P}(\cdot) : A \mapsto [0,1]$
1. Non-negativity	$\mathbb{P}(A) \geq 0, \quad \forall A \subseteq \Omega$
2. Unitarity	$\mathbb{P}(\Omega) = 1$
3. Countable additivity	$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$, where $\{A_1, A_2, A_3, \ldots\}$ is a sequence of **mutually disjoint events**

Probability Properties		**Proof**
Probability of empty set	$\mathbb{P}(\varnothing) = 0$	$\varnothing \cup \varnothing \cup \cdots = \varnothing$ is mutually disjoint set, so by (3), $\mathbb{P}(\varnothing) = \mathbb{P}(\varnothing) + \mathbb{P}(\varnothing) + \cdots + \mathbb{P}(\varnothing)$, which only holds if $\mathbb{P}(\varnothing) = 0$
Finite additivity	$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i)$, where $\{A_1, A_2, A_3, \ldots, A_n\}$ is a *finite* set of **mutually disjoint events**	Let $A_1, \ldots, A_n$ be mutually disjoint, construct $A_1, \ldots, A_n, \varnothing, \varnothing, \ldots$. Since $\mathbb{P}(\varnothing) = 0$ and by countable additivity, $\mathbb{P}(A_1 \cup \cdots \cup A_n \cup \varnothing) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots + \mathbb{P}(A_n) + 0 + 0 \cdots$
Complement rule	$\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$	$\mathbb{P}\left(A \cup A^C\right) = \mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(\Omega) = 1$
Monotonicity	If $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$	If $A \subseteq B$, then $B = A \cup (B \setminus A)$, which is disjoint, thus $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$ since probabilities are non-negative
Probability lies in unit interval	$0 \leq \mathbb{P}(A) \leq 1$	$\varnothing \subseteq A \subseteq \Omega$, thus $\mathbb{P}(\varnothing) \leq \mathbb{P}(A) \leq \mathbb{P}(\Omega) \implies 0 \leq \mathbb{P}(A) \leq 1$
Addition theorem	$\boxed{\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)}$	$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) = \mathbb{P}(A) + \mathbb{P}(B \setminus (A \cap B))$ And $\mathbb{P}(B) = \mathbb{P}(B \setminus (A \cap B)) + \mathbb{P}(A \cap B)$, eliminate $\mathbb{P}(B \setminus (A \cap B))$
Continuity	(i) Let $\{A_n : n \geq 1\}$ be an increasing sequence of events, namely, $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$, and define $A = \bigcup_{n=1}^{\infty} A_n$. Then $\mathbb{P}(A) = \lim_{n\to\infty} A_n$ (ii) Let$\{A_n : n \geq 1\}$ be a decreasing sequence of events, namely, $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$, and define $A = \bigcap_{n=1}^{\infty} A_n$. Then $\mathbb{P}(A) = \lim_{n\to\infty} A_n$	

Classical Probability Model	
Criteria	(i) The sample space Ω is a *finite* set, consisting of N outcomes (ii) The probability of each outcome in Ω occur equally likely. $\mathbb{P}(\{\omega\}) = \frac{1}{N}, \forall \omega \in \Omega$
For any event A	$\mathbb{P}(A) = \frac{\#A}{\#\Omega}$

Conditional probability

Conditional probability	$\boxed{\mathbb{P}(A \mid B) := \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}}$
Multiplication theorem	$\mathbb{P}(A \cap B) = \mathbb{P}(B)\,\mathbb{P}(A \mid B)$
Positive and negative relations	$\begin{cases} \text{Positive relation : } \mathbb{P}(A \mid B) > \mathbb{P}(A) \\ \text{Negative relation : } \mathbb{P}(A \mid B) < \mathbb{P}(A) \end{cases}$

Law of Total Probability and Bayes' theorem

> Intuition for the Law of Total Probability: We think of an event as the effect /result due to one of several distinct causes/reasons. In this way, we compute the probability of the event by *conditioning* on each of the possible causes and adding up all these possibilities. We interpret A as an effect and interpret a partition $\{B_i\}$ as the possible distinct causes leading to the effect A.

Law of Total Probability & Bayes' Theorem

Law of Total Probability	$\boxed{\mathbb{P}(A) = \sum_i \mathbb{P}(A \cap B_i) = \sum_i \mathbb{P}(B_i)\,\mathbb{P}(A \mid B_i)}$, where $\{B_i\}$ is a *partition* of Ω
Bayes' Theorem	$\boxed{\mathbb{P}(A_i \mid B) = \dfrac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \dfrac{\mathbb{P}(B \mid A_i)\,\mathbb{P}(A_i)}{\sum_j \mathbb{P}(A_j)\,\mathbb{P}(B \mid A_j)}}$, for a given partition $\{A_j\}$
Bayes' theorem for the partition $\Omega = A \cup B$	$\mathbb{P}(A \mid B) = \dfrac{\mathbb{P}(B \mid A)\,\mathbb{P}(A)}{\mathbb{P}(B \mid A)\,\mathbb{P}(A) + \mathbb{P}(B \mid A^c)\,\mathbb{P}(A^c)}$
Partition of Ω	The set of events $\{A_i\}$ is a partition of Ω if they are mutually disjoint and collectively exhaustive : $(A_i \cap A_j = \varnothing) \cap (\bigcup_i A_i = \Omega)$

Tree Diagrams

- A tree diagram can be used to represent a conditional probability space
- Each node represents an event, and each edge is weighted by the corresponding probability (the root node is the certain event, with probability of one)
- Each set of sibling branches is a partition of the parent event, and sums to one
- The Law of Total Probability is equivalent to reaching the 'effect' (leaf node) though any of the possible 'causes' (intermediary branches). Multiply along the branches and add the result.

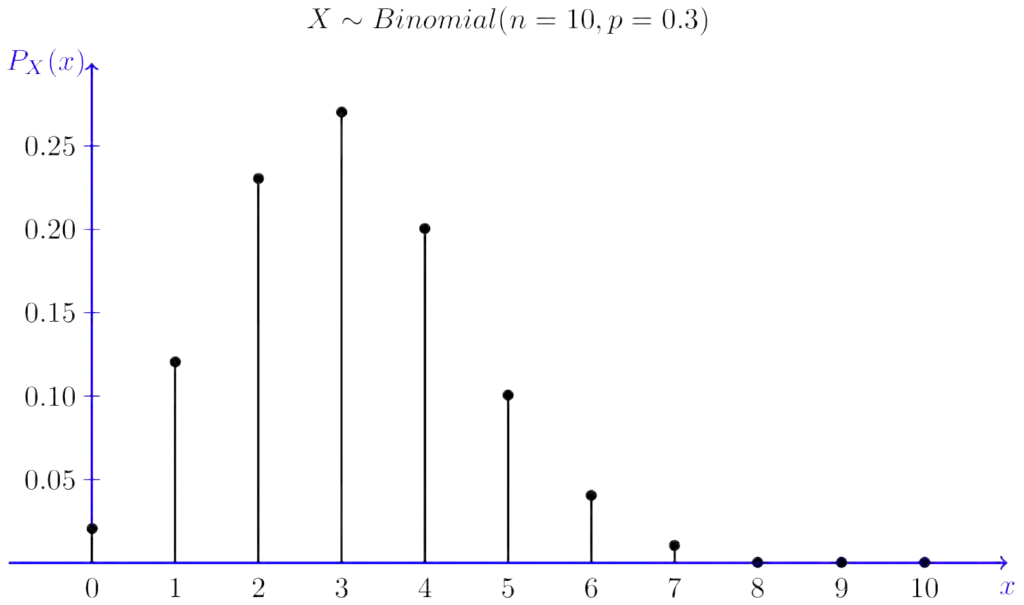
3 Random Variables

Random Variables	
Random Variables	A random variable maps <u>*outcomes*</u> of the sample space to a *real number.* $X : \Omega \to \mathbb{R}, \omega \mapsto X(\omega)$
State space	The state space of a random variable is the range of X, the set of all possible values of $X(\omega)$. $S_X \subseteq \mathbb{R} = \{X(\omega) : \omega \in \Omega\}$
Events of a random variable	*Events* of a random variable for a given $x \in S_X$, are the events in the sample space that are mapped by the random variable to the value x. $A_x = \{\omega \in \Omega : X(\omega) = x\}$
Probability of an event	$\mathbb{P}(X = x) = \mathbb{P}(A_x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\})$
Distribution of a random variable	$(X, \mathbb{P}(X = x))$

Discrete Random Variables

> Where the state space S_X is *countable*

Probability Mass Functions (pmf) p_X	
Discrete random variables	A discrete r.v. is one for which the state space S_X is *countable*
Probability Mass Functions	The probability mass function of X maps the outcomes of X to a probability. $p_X : S_X \to [0,1], p_X(x) = \mathbb{P}(X = x)$
Probability of a given event in the sample space	$\mathbb{P}(x \in A) = \sum_{x \in A} \mathbb{P}(X = x) = \sum_{x \in A} p_X(x)$
Properties of a PMF	1. $p_X(x) \geq 0, \quad \forall x \in S_X$ 2. $\sum_{x \in S_X} p_X(x) = 1$
PMF from CDF	$\mathbb{P}(X = k) = F_X(k) - F_X(k-1)$, for $S_X \in \mathbb{Z}$



Geometric distribution

> The number of **failures** (k or x) before the *first* success in an infinite sequence of independent Bernoulli trials with probability of success p.

Geometric distribution	$X \sim G(p)$
Sample space and random variable state space	$S_X = \{0, 1, 2, 3, \ldots\}$
PMF	$\boxed{\mathbb{P}(X = k) = (1-p)^k p}$, where $k \in \{0, 1, 2, 3, \ldots\}$
CDF	$1 - (1-p)^{k+1}$
Expectation	$\mathbb{E}[X] = \dfrac{1-p}{p}$
Variance	$\text{Var}[X] = \dfrac{1-p}{p^2}$
Memoryless property	$\mathbb{P}(X \geq a + b \vert X \geq b) = \mathbb{P}(X \geq a)$

Derivation: $\mathbb{P}(X = 0) = \mathbb{P}(S) = p$, $\mathbb{P}(X = 1) = \mathbb{P}(FS) = (1-p)p$, $\mathbb{P}(X = 2) = \mathbb{P}(FFS) = (1-p)^2 p$ and so on.

Warning: Careful of the parametrisation of the geometric distribution! *MAST20004 Probability* uses the number of **failures** (X) before the first success, but a common alternate parametrisation is the number of **trials** before the first success ($Y = X + 1$).

Poisson distribution	$X \sim \text{Pn}(\lambda), \quad \lambda > 0$
State space	$S_X = \{0, 1, 2, 3, \ldots\}$
PMF	$\boxed{p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}}$, where $k \in \{0, 1, 2, 3, \ldots\}$
Time-rate of events	If $r =$ time rate of events $[1/\text{time}]$, then $\mathbb{P}\left(\text{k events in interval } t)\right) = \frac{(rt)^k e^{-rt}}{k!}$
Expectation	$\mathbb{E}\left[X\right] = \lambda$
Variance	$\text{Var}[X] = \lambda$
General Poisson approximations	1. $X = X_1 + \cdots + X_m, \quad X_i \sim \text{Bi}(1, p_i)$, where m large, p_i small and need not be equal. 2. $X_1, \ldots, X_m$ are 'weakly dependent' Then, $X \sim \text{Pn}(\lambda = p_1 + \cdots + p_m)$
Poisson approximation to the Binomial	For a binomial $X \sim \text{Bi}(n, p) \approx \text{Pn}(\text{np})$, for p small.
Convolution	$X \sim \text{Pn}(\lambda), Y \sim \text{Pn}(\mu), X \perp Y \implies X + Y \sim \text{Pn}(\lambda + \mu)$

Derivation: We can consider taking the Binomial pmf with $p = \lambda/n$ and shrinking the time period to 0 by taking $n \to \infty$: $\mathbb{P}\left(X = k\right) = \binom{n}{k} (\lambda/n)^x (1 - \lambda/n)^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda}$

Poisson Modelling:

- The occurrence of one event does not affect the probability that a second event will occur. That is, events occur independently.
- The average rate at which events occur is independent of any occurrences. For simplicity, this is usually assumed to be constant, but may in practice vary with time.
- Two events cannot occur at exactly the same instant; instead, at each very small sub-interval exactly one event either occurs or does not occur.

Poisson Approximations

Conditions:

1. The random variable X can be written as the sum of m Bernoulli variables with parameter p_i: $X = X_1 + \cdots + X_m, \quad X_i \sim \text{Bi}(1, p_i)$. Note that the Bernoulli trials need not be fully independent, nor do they need to share the same success probability p.
2. The number of Bernoulli variables (m) is very large, each p_i is very small, and $P_1 + \cdots + P_m$ is of 'normal' scale.
3. The dependence among the random variables $X_1, \ldots, X_m$is weak, in the sense that for each X_i, the number of random variables that are dependent on (related to) X_i is much smaller than m.

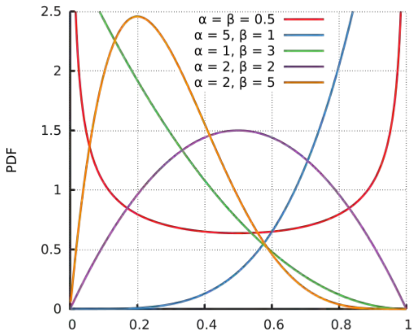
Then, we can approximate the pmf of X by the Poisson distribution with parameter $\lambda = \sum_i^m p_i$.

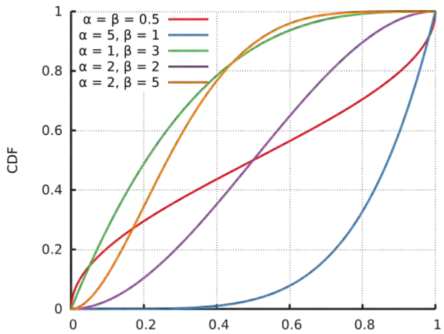
Beta distribution

Beta distribution	$X \sim \text{Beta}(\alpha, \beta), \quad \alpha, \beta > 0$
PDF	$\boxed{f_X(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}}, \quad 0 \leq x \leq 1$
Mean	$\mathbb{E}[X] = \frac{\alpha}{\alpha+\beta}$
Variance	$\text{Var}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Moments	$\mathbb{E}[X^k] = \frac{B(\alpha+k, \beta)}{B(\alpha, \beta)} = \frac{\Gamma(a+k)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta+k)}$

Beta function

Beta function	
Beta function	$\boxed{B(\alpha, \beta) \equiv \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\, dx, \quad \alpha, \beta > 0}$
Beta and Gamma function	$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$





Exponential distribution

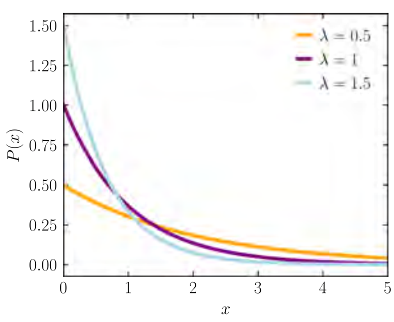
> Consider **infinite** independent Bernoulli trials in **continuous** time with rate α, then X is the **time until the first success**. Exponential distributions are often used to model the waiting time *between* the occurrence of events.
>
> The parameter λ is naturally interpreted as the average number of arrivals per unit interval. More generally, we can always regard α as the rate at which the underlying event occurs per unit time.
>
> The Exponential distribution is the **continuous** time analogue of the **Geometric** (measures time until first success)

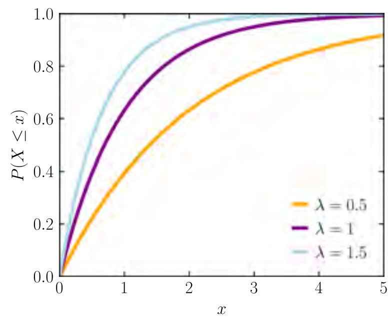
Exponential distribution	$X \sim \exp(\lambda), \quad \lambda > 0$
PDF	$\boxed{f_X(t) = \lambda e^{-\lambda t}, \quad t \geqslant 0}$
CDF	$\boxed{F_X(t) = \mathbb{P}(X \leqslant t) = \begin{cases} 1 - e^{-\lambda t}, & t \geqslant 0 \\ 0, & t < 0 \end{cases}}$
Expectation	$\mathbb{E}[X] = \frac{1}{\lambda}$
Variance	$\mathrm{Var}[X] = \frac{1}{\lambda^2}$
Memoryless property	$\mathbb{P}(X \geq s + t \mid X \geq s) = \mathbb{P}(X \geq t)$
As special case of Gamma Distribution	$\exp(\lambda) \sim \gamma(1, \lambda)$

Note: The geometric distribution is a particular case of the gamma distribution.

Intuition for rate parameter: If you receive phone calls at an average rate of 2 per hour, then you can expect to wait *half an hour* for every call.

Intuition for memoryless property: When X is interpreted as the waiting time for an event to occur relative to some initial time, this relation implies that, if X is conditioned on a failure to observe the event over some initial period of time s, the distribution of the remaining waiting time is the same as the original unconditional distribution. For example, if an event has not occurred after 30 seconds, the conditional probability that occurrence will take at least 10 more seconds is equal to the unconditional probability of observing the event more than 10 seconds after the initial time.





Continuous Bivariate Random Variables

Bivariate Joint PDFs	$f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$
Bivariate Joint PDF	$\boxed{\mathbb{P}\left((X,Y) \in D\right) = \iint_D f_{X,Y}(x,y)\,dx\,dy, \quad D \subseteq \mathbb{R}^2}$
Probability of a rectangular region	$\mathbb{P}(a \leqslant X \leqslant b, c \leqslant Y \leqslant d) = \int_a^b \int_c^d f_{X,Y}(x,y)\,dy\,dx = \int_c^d \int_a^b f_{X,Y}(x,y)\,dx\,dy$
Unitarity	$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$
Joint CDF to joint PDF	$\boxed{f_{X,Y} = \frac{\partial^2}{\partial y \partial x} F_{X,Y} = \frac{\partial^2}{\partial x \partial y} F_{X,Y}}$
Joint PDFs to Marginal PDFs	$\boxed{f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,d\color{red}{y}, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,d\color{red}{x}}$

Warning:

- The joint pdf uniquely determines *the* marginal pdf's, but the converse is not true. Marginal pdf's do not determine the joint distribution.
- To find the marginal PDFs from the Joint PDFs, we integrate the *other* variable out, meaning that our bounds on our integral will be in terms of the *other* variable, and not in terms of the marginal variable.

Recovering joint pdf from joint cdf by partial differentiation

$$F_{X,Y} = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)\,dv\,du$$

Let the inner integral $\varphi(u;y) = \int_{-\infty}^{y} f_{X,Y}(u,v)\,dv$, then $F_{X,Y} = \int_{-\infty}^{x} \varphi(u;y)\,du$. Thus, $\frac{\partial}{\partial x} F_{X,Y}(x,y) = \varphi(x;y) = \int_{-\infty}^{y} f_{X,Y}(x,v)\,dv$, and thus $\frac{\partial^2}{\partial x \partial y} F_{X,Y} = f_{X,Y}$.

Bivariate Joint CDFs	$F_{X,Y} : \mathbb{R}^2 \to [0,1]$
Bivariate Joint CDF	$F_{X,Y}(x,y) = \mathbb{P}\left(X \leq x, Y \leq y\right) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)\,du\,dv$
Non-negativity	$F_{X,Y}(x,y) \geq 0, \quad \forall (x,y) \in \mathbb{R}^2$
Limit to infinity	$\lim_{x,y \to \infty} F_{X,Y}(x,y) = 1$

***General* bivariate normal distributions**	
General bivariate normal distributions	$(X,Y) \sim \mathcal{N}_2\left(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho \in [-1,1]\right)$ where $\left(\frac{X-\mu_X}{\sigma_X}, \frac{Y-\mu_Y}{\sigma_Y}\right) \sim \mathcal{N}_2(\rho)$
Joint pdf	$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2\left(1-\rho^2\right)}\left(\frac{\left(x-\mu_X\right)^2}{\sigma_X^2} - \frac{2\rho\left(x-\mu_X\right)\left(y-\mu_Y\right)}{\sigma_X\sigma_Y} + \frac{\left(y-\mu_Y\right)^2}{\sigma_Y^2}\right)\right)$
Marginal distributions of general $\mathcal{N}_2$	$X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$
Conditional distributions of general $\mathcal{N}_2$	$X\mid_{Y=y} \sim N\left(\rho\sigma_X y_s + \mu_X, \sigma_X^2\left(1-\rho^2\right)\right)$, where $x_s = \frac{x-\mu_X}{\sigma_X}, y_s = \frac{y-\mu_Y}{\sigma_Y}$ $Y\mid_{X=x} \sim N\left(\rho\sigma_Y x_s + \mu_Y, \sigma_Y^2\left(1-\rho^2\right)\right)$
Linear transformation	If (X,Y) is a bivariate normal random variable and $a,b,c,d \in \mathbb{R}$, then $(aX+bY, cX+dY)$ is also a bivariate normal random variable.
Decomposition of the bivariate normal	$\boxed{X = \mu_X + \frac{\sigma_X\rho}{\sigma_Y}(Y-\mu_Y) + \sigma_X\sqrt{1-\rho^2}Z}$
Decomposition of the univariate normal	$\boxed{X = \sigma_X Z + \mu_X}$
Constructing bivariate normals from standard normals Z_1, Z_2	$(X_1 = Z_1,\ X_2 = \rho Z_1 + \sqrt{1-\rho^2}Z_2) \sim \mathcal{N}_2(\rho)$

Geometry of *Standard* Bivariate normal distributions

- The peak of the graph is at the origin $(0,0)$
- The value of $f_{X,Y}(x,y)$ decays to zero in all directions as (x,y) approaches infinity.
- positive relationship if $\rho > 0$, negative relationship if $\rho < 0$
- As $|\rho| \to 1$, the relationship between X and Y becomes stronger.

Marginal distributions from Standard bivariate normal distribution

- It is in general not possible to reconstruct the joint distribution from the marginal distributions.
- We know that the marginal distributions of $N_2(\rho)$ are both standard normal distribution. However, the converse is not true in general – the *joint distribution* of two standard normal random variables needs not be bivariate normal.
- In particular, note that $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(0,1)$ does not imply that $(X,Y) \sim \mathcal{N}_2(\rho)$.

Sufficient condition to determine whether (X,Y) is a bivariate normal random variable:

Conditional Expectation

Conditioning on a random variate	$\boxed{\mathbb{E}[X \mid Y = y] = \begin{cases} \displaystyle\sum_{x \in S_X} x\, p_{X\mid Y}(x \mid y) \\ \displaystyle\int_{-\infty}^{\infty} x\, f_{X\mid Y}(x \mid y)\, dx \end{cases}}$
Conditioning on a random variable	Define $\eta(y) = \mathbb{E}[X \mid Y = y]$ (as above), a number in terms of y. Then $\mathbb{E}[X \mid Y] = \eta(Y)$, the function composed with the random variable Y, itself a random variable
Law of Total Expectation	$\boxed{\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]]}$
Law of Total Expectation (applied)	Multivariate case: $\boxed{\mathbb{E}[\psi(X,Y)] = \mathbb{E}[\mathbb{E}[\psi(X,Y) \mid Y]]}$ Products: $\mathbb{E}[XY] = \mathbb{E}[\mathbb{E}[XY \mid Y]] = \mathbb{E}[Y\,\mathbb{E}[X \mid Y]]$ In terms of probability functions: $\boxed{\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]] = \sum_y \mathbb{E}[X \mid Y = y]\,\mathbb{P}(Y = y)}$
Conditional expectation of an event A	$\boxed{\mathbb{E}[X \mid A] = \begin{cases} \displaystyle\sum_{x \in S_X} x p_{X\mid A}(x) \\ \displaystyle\int_{-\infty}^{\infty} x\, f_{X\mid A}(x)\, dx \end{cases}}$

Conditional Variance

Law of Total Variance	$\boxed{\mathrm{Var}[X] = \mathrm{Var}[\mathbb{E}[X \mid Y]] + \mathbb{E}[\mathrm{Var}[X \mid Y]]}$
Conditional variance of an event A	$\boxed{\mathrm{Var}[X \mid A] = \mathbb{E}\left[X^2 \mid A\right] - (\mathbb{E}[X \mid A])^2}$

Convolutions of independent random variables

Convolutions of *independent* random variables	$\psi(X,Y) = X + Y$
Discrete convolution formula	$\boxed{p_{X+Y}(k) = \sum_{i=0}^{k} p_X(i) p_Y(k-i), \quad k \in \{0,1,2\ldots\}}$ (Where X, Y are independent and non-negative)
Continuous convolution formula	$\boxed{f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx = \int_{-\infty}^{\infty} f_Y(y) f_X(z-y)\, dy}$

Warning: Careful with convolution limits or terminals in the continuous case.

Random Sums using the pgf

Let $\{X_i : 1 \leq i \leq N\}$ be a sequence of N independent identically distributed random variables, where N is an independent non-negative discrete random variable. Let $S_N = \sum_{I=1}^{N} X_i$ be the random sum. Then the generating function of S_N is $G_{S_N}(z = G_N(G_X(z)))$; since
$G_{S_N}(z) = \mathbb{E}\left[z^{S_N}\right] = \mathbb{E}\left[z^{\sum_{i=1}^N X_i}\right] = \mathbb{E}\left[\mathbb{E}\left[z^{\sum_{i=1}^N X_i} \mid N\right]\right] = \mathbb{E}\left[(G_X(z))^N\right] = G_N(G_X(z))$, by definition of G_{S_N}, definition of S_N, law of total expectation, using the convolution theorem, and definition of G_X, G_N, respectively.

Moment generating function (mgf)

> *Arbitrary* random variable. The k-th coefficient of the mgf is the k-th moment of X

Moment generating function (mgf)	$\operatorname{dom} M_X = T = \{t \in \mathbb{R} : \mathbb{E}\left[e^{tX}\right] < \infty\}$
Moment generating function (mgf)	$\boxed{M_X(t) := \mathbb{E}\left[e^{tX}\right] = \begin{cases} \sum_{x \in S_X} e^{tx}\, p_X(x) \\ \int_{-\infty}^{\infty} e^{tx}\, f_X(x)\, dx \end{cases}}$
Uniqueness theorem	Let X, Y be two random variables. Suppose that $M_X(t), M_Y(t)$ are both well defined and equal *in some neighbourhood of the origin* $t = 0$. Then X and Y have the same distribution.
Relation between the pgf and the mgf	$M_X(t) = P_X(e^t)$
Convolution theorem for mgfs	$X \perp Y \implies M_{X+Y}(t) = M_X(t)M_Y(t)$
Linear transformation	$M_{aX+b}(t) = M_X(at)e^{bt}$
Moments via Taylor expansion	$M_X(t) = \sum_{n=0}^{\infty} \mathbb{E}\left[X^n\right] \frac{t^n}{n!}$
Computing *moments*	$\boxed{\mathbb{E}\left[X^n\right] = M_X^{(n)}(0)}$
Computing *central moments*	$\mathbb{E}\left[(X-\mu)^n\right] = \frac{\mathrm{d}^n}{\mathrm{d}t^n}(e^{-\mu t} M_X(t))\big\|_{t=0}$
Expectation	$\mathbb{E}\left[X\right] = M'_X(0)$
Variance	$\operatorname{Var}[X] = M''_X(0) - [M'_X(0)]^2$