

Probability Notes

Probability

- Definition (Probability Function)
 - A probability function \mathbb{P} is some real-valued function on event space $\mathcal{F} = \{A : A \subseteq \Omega\}$, where Ω is a given sample space, that satisfies the Kolmogorov probability axioms:
 - $\forall A \in \mathcal{F} : \mathbb{P}A \geq 0$ (i.e. the probability of any event is non-negative),
 - $\mathbb{P}\Omega = 1$, and
 - $\mathbb{P}\bigcup_{i=1}^m A_i = \sum_{i=1}^m \mathbb{P}A_i$ iff $\forall i \neq j : A_i \cap A_j = \emptyset$ (i.e. iff the events A_1, \dots, A_m are mutually exclusive, then the probability of the event $\bigcup_{i=1}^m A_i$ is equal to the sum of the probabilities of the events A_1, \dots, A_m).
 - Where possible, refer to Ω rather than \mathcal{F} . For example, write $\forall A \subseteq \Omega : \mathbb{P}A \geq 0$ rather than $\forall A \in \mathcal{F} : \mathbb{P}A \geq 0$.
- Common Results
 - $\forall A \in \mathcal{F} : 0 \leq \mathbb{P}A \leq 1$. The probability of any event is no less than 0 and no greater than 1.
 - $\mathbb{P}\emptyset = 0$. The probability of the event \emptyset , which can be interpreted as the event that nothing happens, is 0.
 - $\forall A \in \mathcal{F} : \mathbb{P}A^C = 1 - \mathbb{P}A$ where $A^C = \{a^c : a^c \in \Omega, a^c \notin A\}$. The probability of the complement of some event is equal to 1 minus the probability of that event.
 - $\forall A, B \in \mathcal{F}, A \subseteq B : \mathbb{P}(B \setminus A) = \mathbb{P}B - \mathbb{P}A$ hence $\forall A, B \in \mathcal{F}, A \subseteq B : \mathbb{P}B \geq \mathbb{P}A$. If some event is a subset of another, the probability of the latter is greater than the probability of the former.
 - $\forall A, B \in \mathcal{F} : \mathbb{P}(A \cup B) = \mathbb{P}A + \mathbb{P}B - \mathbb{P}(A \cap B)$ hence $\forall A, B \in \mathcal{F} : \mathbb{P}(A \cup B) \leq \mathbb{P}A + \mathbb{P}B$. The probability of the union of two events is equal to the sum of the probabilities of the two events less the probability of the intersection of the two events.
 - These common results can be given without proof.
- Definition (Conditional Probability)
 - The conditional probability of event A given event B is denoted as $\mathbb{P}(A|B)$, is given by $\frac{\mathbb{P}(A \cap B)}{\mathbb{P}B}$, and is defined iff $\mathbb{P}B \neq 0$.
- Common Results
 - If B_1, \dots, B_m partition Ω , i.e. $\bigcup_{i=1}^m B_i = \Omega$ and $\forall i \neq j : B_i \cap B_j = \emptyset$, then $\forall A \in \mathcal{F} : \mathbb{P}A = \sum_{i=1}^m \mathbb{P}(A|B_i)\mathbb{P}B_i$. This is the law of total probability.
 - $\forall A, B \in \mathcal{F} : \mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}A}{\mathbb{P}B}$. This is Bayes' rule.
 - These common results can be given without proof.
- Definition (Independence)
 - Event A is independent of event B , denoted as $A \perp B$, iff $\mathbb{P}(A \cap B) = \mathbb{P}A\mathbb{P}B$, $\mathbb{P}(A|B) = \mathbb{P}A$, or $\mathbb{P}(B|A) = \mathbb{P}B$. These conditions are equivalent for $\mathbb{P}A, \mathbb{P}B \neq 0$.
 - Any event with zero probability is trivially independent of every other event.

Random Variables

- Definition (Random Variable)
 - A random variable X is a real-valued function on a given sample space Ω that maps each outcome $\omega \in \Omega$ to some real number $X(\omega)$. " $\{X = x\}$ " is understood as an abbreviation of " $\{\omega \in \Omega : X(\omega) = x\}$ ", then $X = x$, $X \leq x$, and $X \in Y$ describe events.
 - Random variables are denoted by uppercase letters, numbers are denoted by lowercase letters.

Univariate Distributions

- Definition (Discrete Random Variable)
 - A random variable X is discrete iff it takes a countable number of distinct values.
- Definition (Support)
 - The support \mathcal{X} of discrete random variable X is the set of values that X takes with non-zero probability.
- Definition (Continuous Random Variable)
 - A random variable X is continuous iff it takes an uncountably infinite number of distinct values.
- Definition (Support)
 - The support \mathcal{X} of continuous random variable X is the set of values x such that $f_X(x) \neq 0$, i.e. X has non-zero probability density.
- Common Results
 - For any continuous random variable X , for all $x \in \mathbb{R}$, $\mathbb{P}\{X = x\} = 0$, hence $\mathbb{P}\{X \leq x\} = \mathbb{P}\{X < x\}$.
- Definition (Cumulative Distribution Function)
 - The cumulative distribution function of random variable X is the function $F_X(x) := \mathbb{P}\{X \leq x\}$.
- Common Results
 - $\forall X : \mathbb{P}\{a \leq X \leq b\} = F_X(b) - F_X(a)$.
 - $\forall X : \forall x_1 \leq x_2 : F_X(x_1) \leq F_X(x_2)$. Every cumulative distribution function is weakly increasing.
 - $\forall X : \lim_{x' \downarrow x} F_X(x') = F(x)$. Every cumulative distribution function is right-continuous.
 - $\forall X : \lim_{x \rightarrow -\infty} F_X(x) = 0, \lim_{x \rightarrow +\infty} F_X(x) = 1$.

- These common results can be given without proof.
- Definition (Probability Mass Function)
 - The probability mass function of discrete random variable X is the function $f_X(x) := \mathbb{P}\{X = x\}$.
- Common Results
 - $\forall X : \forall x : 0 \leq f_X(x) \leq 1$.
 - $\forall X : \sum_{x \in \mathcal{X}} f_X(x) = 1$.
 - $\forall X : F_X(x) = \sum_{u \leq x, u \in \mathcal{X}} f_X(u)$.
- Definition (Probability Density Function)
 - The probability density function of random variable X is the function $f_X(x)$ such that $F_X(x) = \mathbb{P}\{X \leq x\} = \int_{-\infty}^x f_X(u) du$, i.e. such that the cumulative distribution function can be written as an integral of f_X .
- Common Results
 - $\forall X : f_X(x) \geq 0$.
 - $\forall X : \int_{-\infty}^{\infty} f_X(x) dx = 1$.
 - These results are exactly analogous to those of the probability mass function.
 - Note that the probability density function of any random variable has a lower bound at 0 but no upper bound.
 - The cumulative density function of a continuous random variable is continuous.

Joint Distributions

- Definition (Joint Cumulative Distribution Function)
 - The joint cumulative distribution function of two random variables, X, Y is the function $F_{X,Y}(x, y) := \mathbb{P}\{X \leq x, Y \leq y\}$.
 - This definition applies to both discrete and continuous random variables.
 - A n -variate distribution function is defined on \mathbb{R}^n . For completeness, when defining a distribution function, include "0 for all other x " where appropriate.

Discrete Random Variables

- Definition (Joint Probability Mass Function)
 - The joint probability mass function of two discrete random variables X, Y is the function $f_{X,Y}(x, y) := \mathbb{P}\{X = x, Y = y\}$.
- Common Results
 - Results exactly analogous to those of the univariate probability mass function obtain.
 - $\forall X, Y : \forall x, y : 0 \leq f_{X,Y}(x, y) \leq 1$.
 - $\forall X, Y : \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} f_{X,Y}(x, y) = 1$.
 - $\forall X, Y : F_{X,Y}(x, y) = \sum_{u \leq x, u \in \mathcal{X}} \sum_{v \leq y, v \in \mathcal{Y}} f_{X,Y}(u, v)$. The joint cumulative distribution function is recoverable from the joint probability mass function.
- Definition (Marginal Probability Mass Function)
 - The marginal probability mass function of discrete random variable X is simply its (univariate) probability mass function $f_X(x) = \sum_{y \in \mathcal{Y}} f_{X,Y}(x, y)$ where Y is some discrete random variable.
 - The marginal probability mass function of each jointly distributed variable is recoverable from the joint probability mass function (which in turn is recoverable from the joint cumulative distribution function).
- Definition (Marginal Cumulative Distribution Function)
 - The marginal cumulative distribution of discrete random variable X is simply its (univariate) cumulative distribution function $F_X(x) = \sum_{u \leq x, u \in \mathcal{X}} f_X(u)$.
- Definition (Conditional Probability Mass Function)
 - The conditional probability mass function of discrete random variable X given discrete random variable Y is the function $f_{X|Y}(x|y) := \mathbb{P}(X = x | Y = y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$.
 - This fits the intuition about conditional probabilities.
- Definition (Conditional Cumulative Distribution Function)
 - The conditional cumulative distribution function of discrete random variable X given discrete random variable Y is the function $F_{X|Y}(x|y) := \mathbb{P}\{X \leq x | Y = y\} = \sum_{u \leq x, u \in \mathcal{X}} f_{X|Y}(u|y)$.

Continuous Random Variables

- Definition (Joint Probability Density Function)
 - The joint probability density function of two continuous random variables X, Y is the function $f_{X,Y}(x, y)$ such that $F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u, v) dv du$.
 - This definition is exactly analogous to the definition of the joint probability mass function of two discrete random variables.
- Common Results
 - Results analogous to those of the univariate probability density function obtain, except that the probability density at any point has no upper bound, whereas the probability mass at any point has an upper bound at 1.
- Definition (Marginal Probability Density Function)
 - The marginal probability density function of continuous random variable X is simply its (univariate) probability density function $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$ where Y is some continuous random variable.

- This definition is exactly analogous to the definition of the marginal probability mass function of two discrete random variables.
- Definition (Marginal Cumulative Distribution Function)
 - The marginal cumulative distribution of continuous random variable X is simply its (univariate) cumulative distribution function $F_X(x) = \int_{-\infty}^x f_X(u)du$.
- Definition (Conditional Probability Density Function)
 - The conditional probability density function of continuous random variable X given continuous random variable Y is the function $f_{X|Y}(x|y) := \frac{f_{X,Y}(x,y)}{f_Y(y)}$.
 - This definition is exactly analogous to that of the conditional probability mass function. The f functions in this case refer to probability density functions while the f functions in the discrete case refer to probability mass functions.
- Definition (Conditional Cumulative Distribution Function)
 - The conditional cumulative distribution function of continuous random variable X given discrete random variable Y is the function $F_{X|Y}(x|y) := \mathbb{P}\{X \leq x | Y = y\} = \int_{-\infty}^x f_{X|Y}(u|y)du$.
- Definition (Independence)
 - Two random variables, X, Y , are independent, denoted as $X \perp\!\!\!\perp Y$ iff for all x, y , $\mathbb{P}\{X \leq x, Y \leq y\} = \mathbb{P}\{X \leq x\}\mathbb{P}\{Y \leq y\}$. Equivalently, this is iff for all x, y , $F_{X,Y}(x, y) = F_X(x)F_Y(y)$ or for all x, y , $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ (where f denotes a probability mass function in the case of discrete random variables, and a probability density function in the case of continuous random variables).

Moments

Expectation

- Definition (Expectation)
 - The expectation of a random variable X is $\mathbb{E}X := \begin{cases} \sum_{x \in \mathcal{X}} x f_X(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x f_X(x) dx & \text{if } X \text{ is continuous} \end{cases}$
 - The expectation of a random variable X is the first (uncentred) moment of X . The k^{th} (uncentred) moment of X is defined as $\mathbb{E}X^k$. The k^{th} centred moment of X is defined as $\mathbb{E}(X - \mathbb{E}X)^k$.
- Common Results
 - If $X = c$ then $\mathbb{E}X = c$.
 - $\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$. This result is the linearity of expectation.
 - If $\forall i \in \{1, \dots, n\} : \mathbb{E}X_i = \mu$, then $\mathbb{E}(\sum_{i=1}^n X_i) = n\mu$.
 - $\mathbb{E}g(X) = \begin{cases} \sum_{x \in \mathcal{X}} g(x) f_X(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} g(x) f_X(x) dx & \text{if } X \text{ is continuous} \end{cases}$
 - If $\forall x \in \mathbb{R} : g(x) \leq h(x)$, then $\mathbb{E}g(X) \leq \mathbb{E}h(X)$. This result is the monotonicity of expectation.
 - If $X \perp\!\!\!\perp Y$, then for any functions g, h , $\mathbb{E}g(X)h(Y) = \mathbb{E}g(X)\mathbb{E}h(Y)$.
 - These common results can be given without proof.

Variance and Standard Deviation

- Definition (Variance)
 - The variance of a random variable X is $var(X) := \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}[X^2 - 2X\mathbb{E}X + (\mathbb{E}X)^2] = \mathbb{E}X^2 - 2\mathbb{E}X\mathbb{E}X + (\mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2$.
 - The variance of a random variable X is the second centred moment of X .
- Common Results
 - $var(a + X) = var(X)$.
 - $var(aX) = a^2 var(X)$.
 - If X is a discrete random variable, then $var(X) = \sum_{x \in \mathcal{X}} (x - \mathbb{E}X)^2 f_X(x)$.
 - If X is a continuous random variable, then $var(X) = \int_{-\infty}^{\infty} (x - \mathbb{E}X)^2 f_X(x) dx$.
- Definition (Standard Deviation)
 - The standard deviation of a random variable X is $sd(X) := \sqrt{var(X)}$.

Standardisation

- Definition (Standardised Counterpart)
 - The standardised counterpart of random variable X is the random variable $Z := \frac{X - \mu}{\sigma}$, where $\mu := \mathbb{E}X$ and $\sigma := sd(X)$. Then, $\mathbb{E}Z = 0$ and $var(Z) = 1$.
- Definition (Standard Normal Distribution)
 - The standard normal distribution is the random variable $Z \sim N[0, 1]$.
- Definition (Standard Normal Cumulative Distribution Function)
 - The standard normal cumulative distribution function is $\phi(z) := F_Z(z)$.

Covariance and Correlation

- Definition (Covariance)
 - The covariance of two random variables X, Y is $cov(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y) = \mathbb{E}XY - (\mathbb{E}X)(\mathbb{E}Y)$.
 - The covariance of two random variables is a measure of the linear association between the two.

- **Common Results**
 - $cov(X, Y) = cov(Y, X)$. This result is the symmetry of covariance.
 - $cov(X, aY + bZ) = acov(X, Y) + bcov(X, Z)$. This result is the linearity of covariance.
 - $cov(X, c) = 0$.
 - If $X \perp\!\!\!\perp Y$, then $cov(X, Y) = 0$, but the converse is not true.
 - $cov(X, X) = var(X)$.
 - $var(aX + bY) = a^2 var(X) + b^2 var(Y) - 2abcov(X, Y)$.
 - If $\{X_i\}_{i=1}^n$ is a set of independent random variables, and $\forall i \in \{1, \dots, n\} : var(X_i) = \sigma^2$, then $var(\sum_{i=1}^n X_i) = \sum_{i=1}^n var(X_i) = n\sigma^2$.
- **Definition (Correlation)**
 - The correlation of two random variables is $corr(X, Y) := \frac{cov(X, Y)}{\sqrt{var(X)var(Y)}} = \frac{cov(X, Y)}{sd(X)sd(Y)}$.
 - The correlation of two random variables is a measure of the linear association between the two that is invariant to any scaling.

Conditional Expectation

- **Definition (Conditional Expectation)**
 - The conditional expectation of X given $Y = y$ is $\mathbb{E}[X|Y = y] := \begin{cases} \sum_{x \in \mathcal{X}} x f_{X|Y}(x|y) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx & \text{if } X \text{ is continuous} \end{cases}$.
- **Common Results**
 - $\mathbb{E}[aX + bY|Z = z] = a\mathbb{E}[X|Z = z] + b\mathbb{E}[Y|Z = z]$. This result is the linearity of conditional expectation.
 - $\mathbb{E}X = \begin{cases} \sum_{x \in \mathcal{X}} \mathbb{E}[X|Y = y] f_Y(y) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} \mathbb{E}[X|Y = y] f_Y(y) dy & \text{if } X \text{ is continuous} \end{cases}$. This result is the law of iterated expectations.
 - Informally, the conditional expectation of X on $Y = y$ is the best guess of X given $Y = y$, then the law of iterated expectations is the result that "on average", the best guess of X given Y is simply the unconditioned best guess of X .
 - $\mathbb{E}[g(Y)X|Y = y] = \mathbb{E}[g(y)X|Y = y] = g(y)\mathbb{E}[X|Y = y]$.
 - If $X \perp\!\!\!\perp Y$, then $\forall y : \mathbb{E}[X|Y = y] = \mathbb{E}[X]$.
- **Secondary Results**
 - $X\mathbb{E}[Y|X] = \mathbb{E}(XY|X)$. This is intuitive in the right to left direction. Informally, X is "known" within the conditional expectation, and so the expectation is not "over" the X factor of XY , and X can be "brought out" of the conditional expectation.
 - This is the property of "conditioning" and generalises as follows. $\mathbb{E}[g(X)Y|X] = g(X)\mathbb{E}[Y|X]$.
 - $\mathbb{E}[\mathbb{E}(XY|X)] = \mathbb{E}[XY]$.
 - $\mathbb{E}[\mathbb{E}(Y|X)|X] = \mathbb{E}[Y|X]$. Informally, expressing the conditional expectation of (the conditional expectation of Y as a function of X) as a function of X , is simply to express the expectation of Y as a function of X .
- **Definition (Conditional Expectation)**
 - The conditional expectation of X given Y is $\mathbb{E}[X|Y] := \mu(Y)$, where $\mu(y) := \mathbb{E}[X|Y = y]$.
 - This conditional expectation is the "best guess" of X as a function of Y , and is a random variable (whereas the earlier conditional expectation is some value).
- **Common Results**
 - Analogous results obtain for $\mathbb{E}[X|Y]$.
 - $\mathbb{E}[aX + bY + c|Z] = a\mathbb{E}[X|Z] + b\mathbb{E}[Y|Z] + c$. This result is the linearity of conditional expectation.
 - $\mathbb{E}X = \mathbb{E}_Y(\mathbb{E}_X[X|Y])$. This result is the law of iterated expectations.
 - $\mathbb{E}[g(Y)X|Y] = g(Y)\mathbb{E}[X|Y]$.
 - If $X \perp\!\!\!\perp Y$, then $\forall y : \mathbb{E}[X|Y] = \mathbb{E}[X]$.