VIP Refresher: Probabilities and Statistics

Simple | Smart | Speed

SCHOOLS

Introduction to Probability and Combinatorics

r **Sample space** – The set of all possible outcomes of an experiment is known as the sample space of the experiment and is denoted by S.

r **Event** – Any subset E of the sample space is known as an event. That is, an event is a set consisting of possible outcomes of the experiment. If the outcome of the experiment is contained in E, then we say that E has occurred.

r **Axioms of probability** – For each event E, we denote P(E) as the probability of event E occuring. By noting $E_1,...,E_n$ mutually exclusive events, we

(ha⋃vethe)3followingaxioms:n∑n(1)06P(E)61(2)P(S)=1(3)PEi=P(Ei)i=1i=1

r **Permutation** – A permutation is an arrangement of r objects from a pool of n objects, in a given order. The number of such arrangements is given by P(n, r), defined as:

()=n!Pn,r

(n−r)!

r **Combination** – A combination is an arrangement of r objects from a pool of n objects, where the order does not matter. The number of such arrangements is given by C(n, r), defined as:

()=P(n,r)=n!Cn,r

r!r!(n−r)!

Remark: we note that for $0 \leqslant r \leqslant n$*, we have* $P(n,r) \geqslant C(n,r)$*.*

Conditional Probability

r **Bayes' rule** – For events A and B such that P(B) > 0, we have:

()=P(B|A)P(A)PA

|B P(B)

Remark: we have $P(A \cap B) = P(A)P(B|A) = P(A|B)P(B)$.

r **Partition** – Let $\{A_i, i \in [[1,n]]\}$ be such that for all i, $A_i \neq \varnothing$. We say that $\{A_i\}$ is a partition if we have:

⋃n∀i6=j,Ai∩Aj=∅andAi=Si=1

∑n*Remark:foranyeventBinthesamplespace,wehave*P(B)=P(B|Ai)P(Ai).i=1

r **Extended form of Bayes' rule** – Let $\{A_i, i \in [[1,n]]\}$ be a partition of the sample space. We have:

()=P(B|A)P(A)PA

|B

∑kkknP(B|Ai)P(Ai)i=1

r **Independence** – Two events A and B are independent if and only if we have:

$P(A \cap B) = P(A)P(B)$

Random Variables

r **Random variable** – A random variable, often noted X, is a function that maps every element in a sample space to a real line.

r **Cumulative distribution function (CDF)** – The cumulative distribution function F, which is monotonically non-decreasing and is such that $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$, is defined as:

F(x)=P(X6x)

Remark: we have $P(a < X \leqslant B) = F(b) - F(a)$.

r **Probability density function (PDF)** – The probability density function f is the probability that X takes on values between two adjacent realizations of the random variable.

r **Relationships involving the PDF and CDF** – Here are the important properties to know in the discrete (D) and the continuous (C) cases.

Case C

∑DFFPDFfPropertieso∑fPDF(D)F(x)=P(X=xi)f(xj)=P(X=xj)06f(xj)61andf(xj)=1xi6xj

ˆx ˆ(dF +∞C)F(x)=f(y)dyf(x)=f(x)>0andf(x)dx=1

−∞ dx −∞

r **Variance** – The variance of a random variable, often noted Var(X) or σ^2, is a measure of the spread of its distribution function. It is determined as follows:

$Var(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2$

r **Standard deviation** – The standard deviation of a random variable, often noted σ, is a measure of the spread of its distribution function which is compatible with the units of the actual random variable. It is determined as f

√ollows:σ=Var(X)

r **Expectation and Moments of the Distribution** – Here are the expressions of the expected value E[X], generalized expected value E[g(X)], kth moment E[Xk] and characteristic function ψ(ω) for the discrete and continuous cases:

Case E[X] E[g(X)] E[Xk

∑∑∑]∑ψ(ω)nnnn(D)xf(x)g(x)f(x)xkf(x)

()iωxiiiiiiifxiei=1i=1i=1i=1

ˆ+∞ˆ+∞ˆ+∞ˆ+∞(C)xf(x)dxg(x)f(x)dxxkf(x)dxf(x)eiωxdx
−∞ −∞ −∞ −∞

Remark: we have eiωx = cos(ωx) + i sin(ωx).

r **Revisiting the** kth **moment** – The kth moment can also be computed with the characteristic function as follows:

[][]=1∂kkψEXik∂ωkω=0

r **Transformation of random variables** – Let the variables X and Y be linked by some function. By noting fX and fY the distribution func

| | tionofXandYrespectively,wehave:()=()dxfYyfXx

| | | | dyr**Leibnizintegralr**

(**ule**–

Letgbeafunctionofxandpotentiaˆllyc,anda,bboundariesthatmaydependonc.

Wehavˆe:)∂b()=∂b∂ab∂ggxdx·g(b)−·g(a)+(x)dx∂ca∂c∂ca∂c

r **Chebyshev's inequality** – Let X be a random variable with expected value μ and standard deviation σ. For k,σ > 0, we have the following inequality:

P(
|X−μ|>kσ)]6k2

Jointly Distributed Random Variables

r **Conditional density** – The conditional density of X with respect to Y , often noted f X| Y , is defined as follows:

()=fXY(x,y)fX
|Yx fY(y)

r **Independence** – Two random variables X and Y are said to be independent if we have:

fXY (x,y) = fX(x)fY (y)

r **Marginal density and cumulative distribution** – From the joint density probability function fXY , we have:

Case Margi

∑naldensityCumul∑ative∑function(D)fX(xi)=fXY(xi,yj)FXY(x,y)=fXY(xi,yj)jxi6xyj6y

ˆ+∞ˆxˆy(C)fxfx,ydy''''X()=XY()FXY(x,y)=fXY(x,y)dxdy
−∞ −∞ −∞

r **Distribution of a sum of independent random variables** – Let Y = X1 + ... + Xn with X1, ..., Xn independent. We have:

∏nψY(ω)=ψX(ω)kk=1

r **Covariance** – We define the covariance of two random variables X and Y , that we note σ2 XY or more commonly Cov(X,Y), as follows:

Cov(X,Y) , σ2
XY =E[(X − μX)(Y − μY)] = E[XY] − μXμY

r **Correlation** – By noting σ X, σY the standard deviations of X and Y , we define the correlation between the random variables X and Y , noted ρXY , as follows:

σ2
ρ XY X Y =
σXσY

Remarks: For any X,Y *, we have* ρXY ∈ [−1,1]. *If X and Y are independent, then* ρXY = 0.

r **Main distributions** – Here are the main distributions to have in mind:

TypeDistribution PD

(**F**)ψ(ω)E[X]Var(X)nX B(n,p)P(X=x)=pxqn−x(peiω+q)nnpnpqxBinomialx∈[[0,n]]

(D)
μx
X Po(μ) P(X=x)= e−μ eμ(eiω−1) μ μ
Poisson x!x
∈ N

e i ω b
()()=1 −eiωaa+b(b−a)2X
U a, b fx b
−aUniform[]

()(b−a)iω212x∈a,b(C)()()=1−1x−μ2X Nμ,σfx√e2eiωμ1ω2σ2σ−2μσ22πσGaussianx
∈ R
X
−iω λ λ2 Exp(λ)f(x)=λe−λx 1 1 11

Parameter estimation

r **Random sample** – A random sample is a collection of n random variables $X_1,...,X_n$ that are independent and identically distributed with X.

r **Estimator** – An estimator $\hat{\theta}$ is a function of the data that is used to infer the value of an unknown parameter θ in a statistical model.

r **Bias** – The bias of an estimator $\hat{\theta}$ is defined as being the difference between the expected value of the distribution of $\hat{\theta}$ and the true value, i.e.:

$\text{Bias}(\hat{\theta})=E[\hat{\theta}]$

$-\ \theta$

Remark: an estimator is said to be unbiased when we have $E[\hat{\theta}] = \theta$.

r **Sample mean and variance** – The sample mean and the sample variance of a random sample are used to estimate the true mean μ and the true variance σ^2 of a distribution, are noted $\overline{X}$ and s^2 respectively

∑,andaresuchthat:n n1X=Xands2=σ̂2i =1(Xi−X)2n n−1i=1 i=1

r**CentralLimitTheorem**–Letushavearandomsample followingagiven distribution with mean μ

σ)

(X1,...,Xnandvariance2,thenwehave: X̄ μ σ √n N√n→+∞