Group 1: Sets and Counting

Set theory

Definition : ☐ set ☐ is a collection of elements

Given 2 sets A and B, we have:

Union $A \cup B$.

Intersection $A \cap B$.

Complement A^c or $S - A$.

Difference $A - B$.

De Morgan's laws

$$(A \cup B)^c = A^c \cap B^c$$

$$(A \cap B)^c = A^c \cup B^c$$

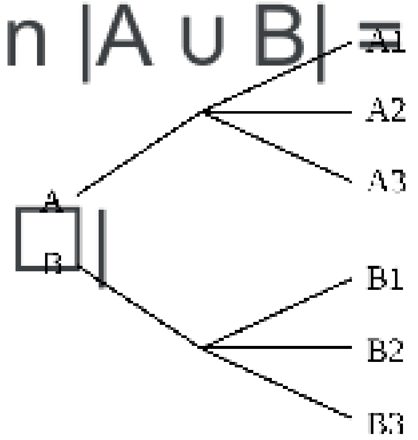
Counting

Rule of sum:

(☐ ∩ ☐ = ∅), then |A ∪ B| = |A| + |B|

Rule of product:

|☐ × ☐| = |☐| · |☐|



Permutations

$$n! = n \times (n-1) \times (n-2) \times \ldots \times 1$$

K-permutation - arrange K out of N objects

$$P_{n,k} = \frac{n!}{(n-k)!}$$

Combinations

$\underbrace{\text{order does not matter}}$

$$C(N,K) = \binom{N}{K} = P(N,K)/K! = \frac{N!}{K!\,(N-K)!}$$

Complementary counting:

$$\underbrace{|S|} = |U| - \underbrace{|S^C|}$$

Group 2: Probability Basics

Venn diagram:

Sample Space	Set Ω of all possible outcome
Event	$E \subseteq \Omega$
Mutual Exclusion	□∩□=∅.

Assume Ω have equally likely elements:
□(□)=|□|/|Ω|

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

axioms of probs
+ non-negative □(□)≥0
+ normalization □(Ω)=1
+ countable activity □(□∪□)=□(□)+□(□) $\rightarrow E \cap F = \phi$

Corrolaries
+ complementation □(□^C)=1−□(□)
+ monotonicity □⊆□ -> □(□)≤□(□)
+ inclusion exclusion □(□∪□)=□(□)+□(□)−□(□

- a **probability measure P: E → [0,1]** with P[Ω]=1 and $P[\cup_i A_i] = \Sigma_i P[A_i]$ for countably many, pairwise disjoint A_i

Properties of P:

$P[A] + P[\neg A] = 1$

$P[A \cup B] = P[A] + P[B] - P[A \cap B]$

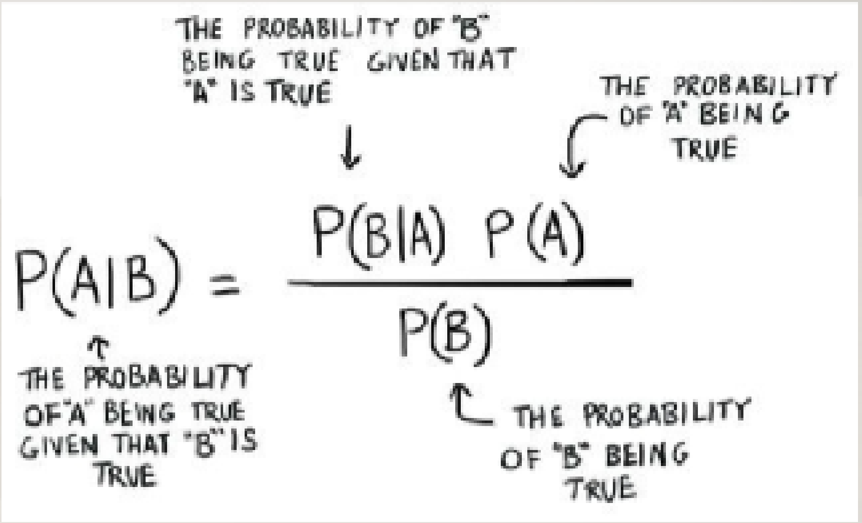
$P[\varnothing] = 0$ (null/impossible event)

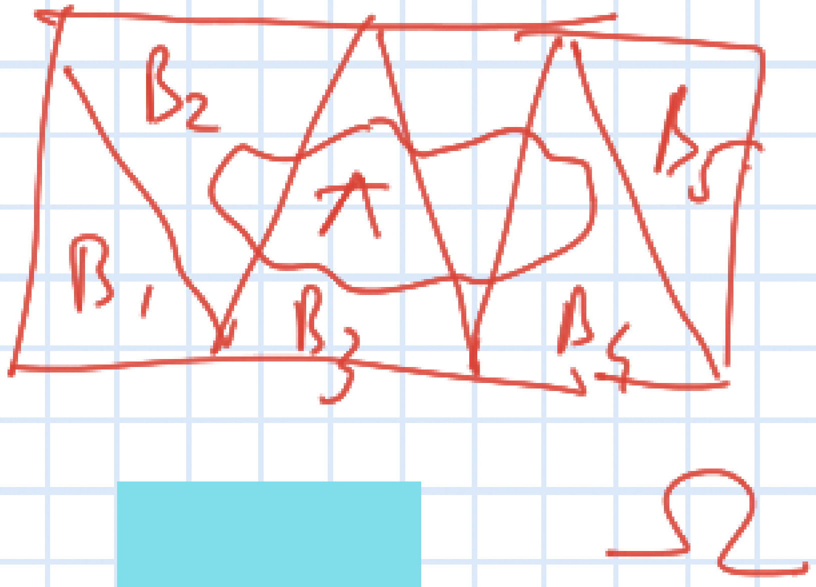
$P[\Omega] = 1$ (true/certain event)

Group 3: Conditional Probability; Independence

1) **Bayes Theorem**

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





$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

2) **Chain Rule**

Prove that

$$P(A_1 \cap A_2 \cap \cdots \cap A_N) = P(A_1) \times P(A_2 \mid A_1) \times P(A_3 \mid A_2 \cap A_1) \times \ldots \times P(A_N \mid A_{N-1} \cap A_{N-2} \cdots \cap A_1)$$

3) **Conditional Independence**

Events A and B are **conditionally independent** given an event C

$$P(A, B|C) = P(A|C)P(B|C)$$

4) **Law of Total Probability (LOTP)**

$$P(A) = \sum_n P(A \mid B_n)P(B_n),$$

$$P(A \mid C) = \sum_n P(A \mid C \cap B_n)P(B_n \mid C)$$

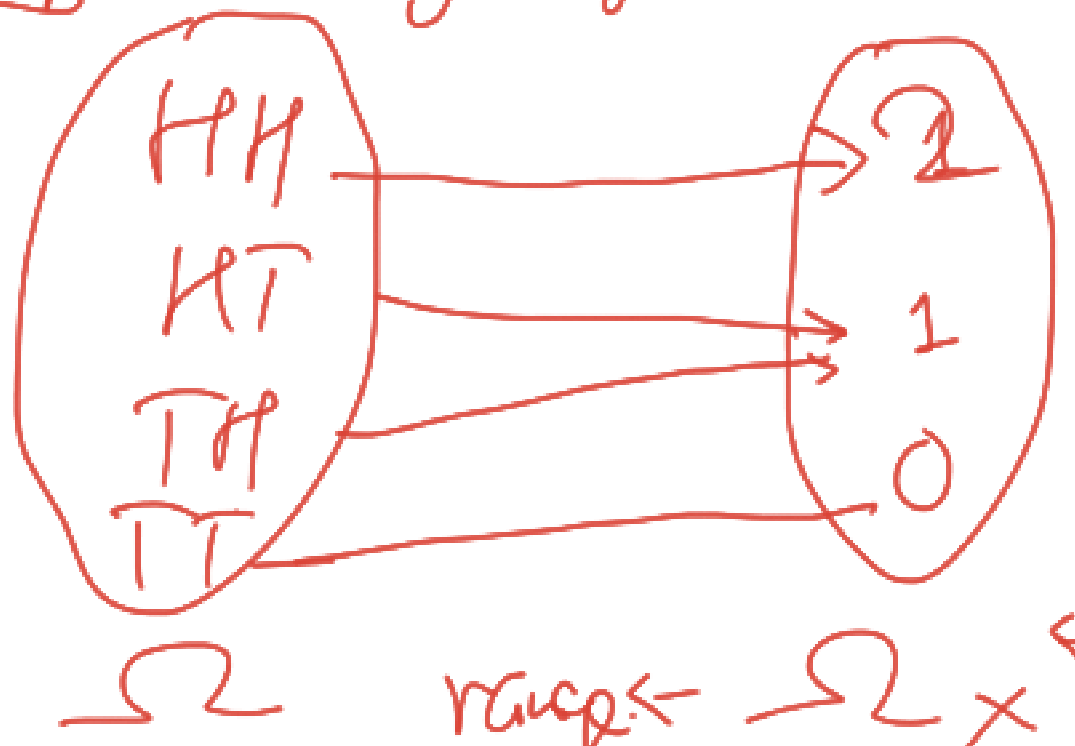
Let $A_1, A_2, \ldots, A_n$ be a partition of sample space Ω, with $P(A_i) > 0$ for all i.

Then,

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i) = P(B|A_1)P(A_1) + \ldots + P(B|A_n)P(A_n)$$

Group 4: Discrete Random Variables Basics

Exp: flipping a fair coin twice



range ← Ω_X

X = # heads

Suppose we conduct an experiment with sample sapce Ω.
A **random variable (rv)** is a numeric function of the outcome:

$$X : \Omega \to R$$
$$\omega \to X(\omega)$$

The set of possible values X can take on is its **range/support**, denoted as: Ω_X

If Ω_X is finite or countably infinite, X is a **discrete** random variable (drv).
If Ω_X is uncountably large, X is a **continuous** random variable (crv).

Expectation

$$E[X] = \sum_{k \in \Omega_X} k . p_X(k)$$

average of the possible values, weighted by their probabilities

PMF

$$p_X : \Omega_X \to [0,1] \qquad \sum_{k \in \Omega_X} p_X(k) = 1$$
$$k \to P(X = k)$$

assign probabilities to possible values of X

CDF

$$F_X(a) = P(X \le a)$$

$$P(a < X \le b) = F_X(b) - F_X(a)$$

$$0 \le F_X(a) \le 1$$

$$\lim_{a \to \infty} F_X(a) = 1 \text{ and } \lim_{a \to -\infty} F_X(a) = 0$$

Group 5: Expectation and Variance of Discrete RVs

Expectation

the weighted average

The **expectation** of a discrete random variable X is:

$$E[X] = \sum_{k \in \Omega_X} k . p_X(k)$$

i.e., we take an average of the possible values, weighted by their probabilities.

Linearity of expectation

the expected value of a sum is the sum of the expected values

Let Ω be the sample space of an experiment, $X, Y : \Omega \rightarrow R$ be random variables both defined on Ω, and $a, b, c \in R$ be scalars. Then,

$$E[X + Y] = E[X] + E[Y]$$

and

$$E[aX + b] = aE[X] + b$$

Combining them gives,

$$E[aX + bY + c] = aE[X] + bE[Y] + c$$

LOTUS

the expectation of a function of a r.v

Let X be a discrete random variable with range Ω_X and $g : D \rightarrow R$ be a function defined at least over Ω_X, $(\Omega_X \subseteq D)$. Then:

$$E[g(X)] = \sum_{b \in \Omega_X} g(b) p_X(b)$$

Note that in general, $E[g(X)] \neq g(E[X])$.
For example, $E[X^2] \neq (E[X])^2$, and $E[\log(X)] \neq \log(E[X])$.

Variance

how spread out the distribution is

The **variance** of a random variable X is defined to be:

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

The **variance** is always nonnegative since we take the expectation of a nonnegative random variable $(X - E[X])^2$. The first equality is the definition of variance, and the second equality is a more useful identity for doing computation.

Standard deviation

variance with extra steps

Another measure of a random variable X's spread is the standard deviation, which is:

$$\sigma_X = \sqrt{Var(X)}$$

This measure is also useful, because the units of variance are squared in terms of the original variable X, and this essentially "undoes" our squaring, returning our units to the same as X.

Property of variance

We can also show that for any scalar $a, b \in R$,

$$Var(aX + b) = a^2 Var(X)$$

Group 6: Gallery of Discrete Random Variables

Independence

Random variables X and Y are independent, denoted $X \perp Y$, if for all $x \in \Omega_X$ and all $y \in \Omega_Y$, any of the following three equivalent properties holds:

- $P(X = x \mid Y = y) = P(X = x)$
- $P(Y = y \mid X = x) = P(Y = y)$
- $P(X = x \cap Y = y) = P(X = x) \cdot P(Y = y)$

Note, that this is the same as the event definition of independence, but it must hold for all events $\{X = x\}$ and $\{Y = y\}$.

Bernoulli

A random variable X is Bernoulli (or indicator), denoted $X \sim Ber(p)$, if and only if X has $\Omega_X = \{0, 1\}$ and the following PMF:

$$p_X(k) = \begin{cases} p & if\ k = 1 \\ 1-p & if\ k = 0 \end{cases}$$

Additionally,

$$E[X] = p \text{ and } Var(X) = p(1-p)$$

"Bernoulli trial": 1 experiment with two outcomes: success or failure

ex: every yes or no question

Binomial

A random variable X has a Binomial distribution, denoted $X \sim Bin(n, p)$, if and only if X has the following PMF for $k \in \Omega_X = \{0, 1, 2, \ldots, n\}$:

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

X is the sum of n independent $Ber(p)$ random variables, and represents the number of "successes" in n independent Bernoulli trails where

$$P(success) = p.$$

Additionally,

$$E[X] = np \text{ and } Var(X) = np(1-p)$$

"given n Bernoulli trials, how many successes?"

ex: coin is tossed n times, what is the probability that heads comes up exactly k times?

n time of Ber(p)

Geometric

X is a Geometric random variable, denoted $X \sim Geo(p)$, if and only if X has the following probability mass function (and range $\Omega_X = \{1, 2, \ldots\}$):

$$p_X(k) = (1-p)^{k-1} p, \quad k = 1, 2, 3, \ldots$$

Additionally,

$$E[X] = \frac{1}{p} \text{ and } Var(X) = \frac{1-p}{p^2}$$

"how many trials until first success?"

ex: coin is repeatedly tossed, what's the probability that the first time heads comes up occurs on the 8th toss?

Negative Binomial

X is a negative binomial random variable, denoted $X \sim NegBin(r, p)$, if and only if X has the following probability mass function (and range $\Omega_X = \{r, r+1, \ldots\}$):

$$p_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k = r, r+1, \ldots$$

X is the sum of r independent $Geo(p)$ random variables.
Additionally,

$$\cancel{E[X] = rp} \text{ and } Var(X) = \frac{r(1-p)}{p^2}$$

"how many trials until r successes?" - ex: 3rd head on the 8th flips

Sum of r Geo(p) vars

$E[x] = \frac{r}{p}$

Uniform

X is a uniform random variable, denoted $X \sim Unif(a, b)$, where a < b are integers, if and only if X has the following probability mass function

$$p_X(k) = \begin{cases} \frac{1}{b-a+1}, & k \in \{a, a+1, \ldots, b\} \\ 0, & otherwise \end{cases}$$

X is equally likely to take on any value in $\Omega_X = \{a, a+1, \ldots, b\}$. This set contains $b - a + 1$ integers, which is why $P(X = k)$ is always $\frac{1}{b-a+1}$.

Additionally,

$$E[X] = \frac{a+b}{2} \text{ and } Var(X) = \frac{(b-a)}{}$$

known, finite number of outcomes equally likely to happen

Poisson

$X \sim Poi(\lambda)$ if and only if X has the following probability mass function (and range $\Omega X = \{0, 1, 2, \ldots\}$):

$$p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots$$

If λ is the historical average of events per unit of time, then X is the number of events that occur in a unit of time.

Additionally,

$$E[X] = \lambda, \ Var(X) = \lambda$$

Number of events that occur in an interval of time

ex: How many babies are born in 1 hour?

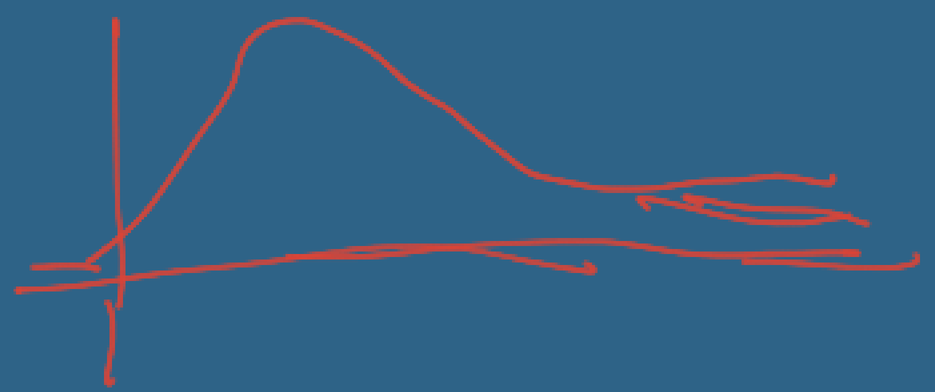
Group 7: Continuous Random Variable Basics

Difference between Discrete and Continuous field

DISCRETE	CONTINUOUS
Random variable X	Random variable X
Probability Mass Function (PMF) $p_{X(k)} = P(X = k)$	Probability Distribution Function (PDF) $f_X(x), k \in \Omega_X$ $P(c \le X \le d) = \int_c^d f(x)dx$
Cumulative Distribution Function (CDF) $F_X(t) = P(X \le t)$	Cumulative Distribution Function (CDF) $F_X(t) = P(X \le t)$

LOTUS: Law of unconscious statistician

	Discrete	Continuous
Expectation/LOTUS	$\mathbb{E}[g(X)] = \sum_x g(x)p_X(x)$	$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx$



PDF

Let X be a continuous random variable. The probability density function (PDF) of X is the function $f_X: \mathbb{R} \to \mathbb{R}$, such that the following properties hold:

- $f_X(t) \ge 0, \forall t \in \mathbb{R}$
- $\int_{-\infty}^{\infty} f_X(t)\,dt = 1$
- $P(a \le X \le b) = \int_a^b f_X(t)\,dt$
- $P(X = y) = 0 \;\forall\, y \in \mathbb{R}$
- The probability that $X \sim q$ is proportional to its density $f_X(q)$;

$$P(X \sim q) = P\left(q - \frac{\epsilon}{2} \le X \le q + \frac{\epsilon}{2}\right) \sim \epsilon f_X(q)$$

CDF

Let X be a continuous random variable. The cumulative distribution function (CDF) of X is the function $F_X: \mathbb{R} \to \mathbb{R}$, such that:

- $F_X(t) = P(X \le t) = \int_{-\infty}^{t} f_X(w)\,dw, \forall w \in \mathbb{R}$
- $\frac{dF_X(u)}{du} = f_X(u)$
- $P(a \le X \le b) = F_X(b) - F_X(a)$
- F_X is monotone increasing, that is $F_X(c) \le F_X(d)$ for $c \le$ d
- $\lim_{v \to -\infty} F_X(v) = 0$
- $\lim_{v \to \infty} F_X(v) = 1$

Group 8: Gallery of Continuous Random Variables

uniform distribution

THE (CONTINUOUS) UNIFORM RV

Uniform (Continuous) RV: $X \sim Unif(a, b)$ where $a < b$ are real number, if and only if X has the following pdf:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & otherwise \end{cases}$$

X is equally likely to take any value in $[a, b]$.

$$E[X] = \frac{a+b}{2} \qquad Var(X) = \frac{(b-a)^2}{2} \qquad F_X(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & \le x \le b \\ 1, & x > b \end{cases}$$

VINUNIVERSITY VINUNIVERSITY

exponential distribution

THE EXPONENTIAL RANDOM VARIABLE

- λ is the average number of events in 1 unit of time
- Range: $[0, \infty)$.
- Notation: Exponential(μ) or exp(μ)
- Density:

$$f_X(x) = \begin{cases} \mu e^{-\mu x} & x \ge 0 \\ 0 & otherwise \end{cases}$$

$$E[X] = \frac{1}{\mu} \qquad Var(X) = \frac{1}{\mu^2} \qquad F_X(x) = \begin{cases} 1 - e^{-\mu x}, & x \ge 0 \\ 0, & otherwise \end{cases}$$

VINUNIVERSITY VINUNIVERSITY

 # hours between 2 car arrivals

poisson process

- **Poisson Distribution:**

  Number of events that occur in an interval of time

- **Exponential Distribution:**

  The time taken between two events occurring

 # cars passing a tollgate in 1 hour

normal distribution

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ if and only if X has the following PDF (and range $\Omega_X = (-\infty, \infty)$):

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

This Normal random variable actually has as parameters its mean and variance, and hence:

$$E[X] = \mu$$
$$Var(X) = \sigma^2$$

The "standard normal" random variable is typically denoted Z has mean 0 and variance 1.

By the closure property of normal, if $Z \sim \mathcal{N}(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

The CDF has no closed-form, but we denote the CDF of the standard normal by

$$\Phi(a) = F_Z(a) = P(Z \le a)$$

Note that by symmetry property:

$$\Phi(-a) = 1 - \Phi(a)$$

standardizing RVs.

Memorylessness

MEMORYLESSNESS

A random variable X is **memoryless** is for all $s, t \ge 0$,

$$P(X > s + t \mid X > s) = P(X > t)$$

Suppose that the probability that a taxi arrives within the first five minutes is p. If I wait five minutes and in fact no taxi arrives, then the probability that a taxi arrives within the next five minutes is still p.

VINUNIVERSITY VINUNIVERSITY

Group 9: Normal Distributions; Standardizing Random Variables

(1)

Standardizing RVs

Let X be random variable (discrete or continuous) with $E[X] = \mu$ and $Var(X) = \sigma^2$.

We call:

$$\frac{X - \mu}{\sigma}$$

a standardized version of X, which measures *how many standard deviations above/below the mean* a point is.

$$E\left[\frac{X - \mu}{\sigma}\right] = \frac{1}{\sigma}(E[X] - \mu) = 0$$

$$Var\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma^2} Var(X - \mu) = \frac{1}{\sigma^2} Var(X) = \frac{1}{\sigma^2}\sigma^2 = 1$$

NRV

X~N(mean = muy, var = sigma^2)

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ if and only if X has the following PDF (and range $\Omega_X = (-\infty, \infty)$):

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

This Normal random variable actually has as parameters its mean and variance, and hence:

$$E[X] = \mu$$
$$Var(X) = \sigma^2$$

(2)

Closure properties of the NRV

Closure under Scale and Shift

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$aX + b \sim \mathcal{N}(a\mu + b, \mu^2\sigma^2)$$

Closure under Addition

If $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and Y $\sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, then

$$aX + bY + c \sim \mathcal{N}(a\mu_X + b\mu_Y + c, a^2\sigma_X^2 + b^2\sigma_Y^2)$$

$\mu = 0, \sigma^2 = 1$

(3)

Standard normal CDF

CDF: F(a) = P(X <= a) = P((X - mean) / std <= (a-mean)/std) = phi((a-mean)/std)

P(a<=X<=b)=P((a-mean)/std <=Z <= (b-mean)/std)=phi(b)-phi(a)

normal dis: đối xứng qua trục x = muy

phi(-a) = 1 - phi(a)

ϕ = table

(4) R($\boxed{\phi(-5)} \simeq 0$

Group 10: MGFs, Law of Large Numbers, and CLT

Things to note

CLT: for a large number (>= 25~30), the distribution can be approximated as a normal distribution

the Mean is the total mean (regardless of independence) , the variance is also the sum of the variances (if the events are independent).

(Uniqueness) The following are equivalent:

a) X and Y have the same distribution

b) $f_X(z) = f_Y(z)$ for all $z \in \mathbb{R}$

c) $F_X(z) = F_Y(z)$ for all $z \in \mathbb{R}$

d) There is an $\epsilon > 0$ such that $M_X(t) = M_Y(t)$ for all $t \in (-\epsilon, \epsilon)$

CLT

Moment Generating Function

$$M_X(t) = E[e^{tX}]$$

Generating Moments with MGFs

$$M'_X(0) = E[X],$$

$$M''_X(0) = E[X^2],$$

and in general, $M_X^{(n)}(0) = E[X^n]$

Derived MGF for some typical distribution

Table 7.2 Continuous Probability Distribution.

	Probability density function, $f(x)$	Moment generating function, $M(t)$	Mean	Variance
Uniform over (a, b)	$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$	$\frac{e^{tb} - e^{ta}}{t(b-a)}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential with parameter $\lambda > 0$	$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$	$\frac{\lambda}{\lambda - t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma with parameters $(s, \lambda), \lambda > 0$	$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{s-1}}{\Gamma(s)} & x \geq 0 \\ 0 & x < 0 \end{cases}$	$\left(\frac{\lambda}{\lambda - t}\right)^s$	$\frac{s}{\lambda}$	$\frac{s}{\lambda^2}$
Normal with parameters (μ, σ^2)	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \quad -\infty < x < \infty$	$\exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$	μ	σ^2

Table 7.1 Discrete Probability Distribution.

	Probability mass function, $p(x)$	Moment generating function, $M(t)$	Mean	Variance
Binomial with parameters n, p; $0 \leq p \leq 1$	$\binom{n}{x} p^x (1-p)^{n-x}$ $x = 0, 1, \ldots, n$	$(pe^t + 1 - p)^n$	np	$np(1-p)$
Poisson with parameter $\lambda > 0$	$e^{-\lambda} \frac{\lambda^x}{x!}$ $x = 0, 1, 2, \ldots$	$\exp\{\lambda(e^t - 1)\}$	λ	λ
Geometric with parameter $0 \leq p \leq 1$	$p(1-p)^{x-1}$ $x = 1, 2, \ldots$	$\frac{pe^t}{1 - (1-p)e^t}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative binomial with parameters r, p; $0 \leq p \leq 1$	$\binom{n-1}{r-1} p^r (1-p)^{n-r}$ $n = r, r+1, \ldots$	$\left[\frac{pe^t}{1 - (1-p)e^t}\right]^r$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$

Group 11: Joint discrete/continuous Random Variables

Let X, Y be discrete random variables.
The joint PMF of X and Y is:

$$p_{X,Y}(a,b) = P(X = a, Y = b)$$

The joint range is the set of pairs (c,d) that have nonzero probabilities:

$$\Omega_{X,Y} = \{(c,d): p_{X,Y}(c,d) > 0\} \subseteq \Omega_X \times \Omega_Y$$

Note that the probabilities must sum to 1

$$\sum_{(s,t)\in\Omega_{X,Y}} p_{X,Y}(s,t) = 1$$

Furthermore, note that if $g: \mathbb{R}^2 \to \mathbb{R}$, then LOTUS extends to the multidimensional case:

$$E[g(X,Y)] = \sum_{x\in\Omega_X}\sum_{y\in\Omega_Y} g(x,y)p_{X,Y}(x,y)$$

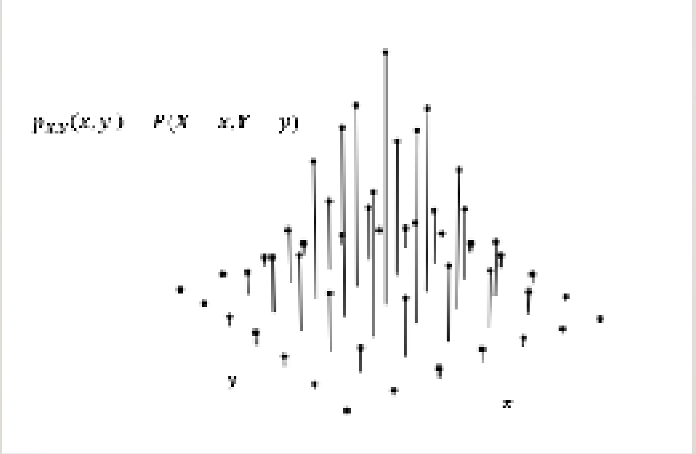
Let X, Y be discrete random variables. The marginal PMF of X is:

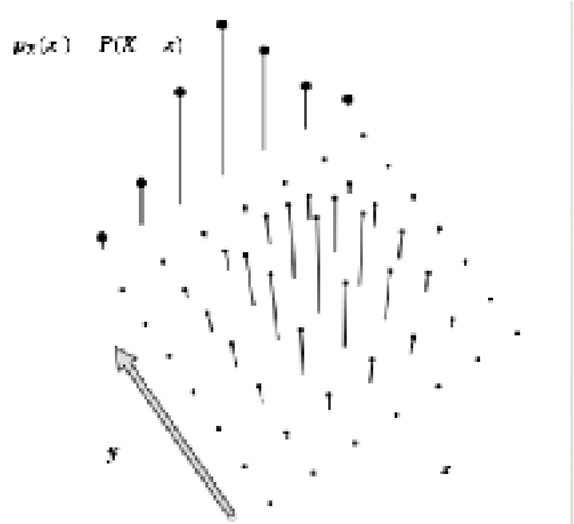
$$p_X(a) = \sum_{b\in\Omega_Y} p_{X,Y}(a,b)$$

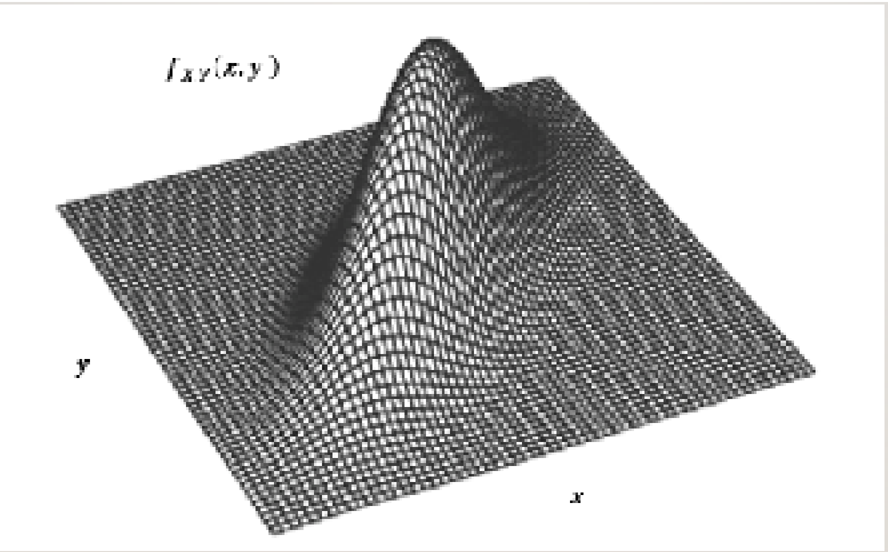
The marginal PMF of Y is:

$$p_Y(b) = \sum_{a\in\Omega_X} p_{X,Y}(a,b)$$

Discrete







Continuos

① Joint PMF
PDF
② marginal
PDF
PMF

Let X, Y be continuous random variables. The joint PDF of X and Y is:

$$f_{X,Y}(a,b) > 0$$

The joint range is the set of pairs (c,d) that have nonzero density:

$$\Omega_{X,Y} = \{(c,d): f_{X,Y}(c,d) > 0\} \subset \Omega_X \times \Omega_Y$$

Note that the double integral over all values must be 1:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$$

Let X, Y be discrete random variables. The marginal PDF of X is:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$$

The marginal PMF of Y is:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$$

Furthermore, note that if $g: \mathbb{R}^2 \to \mathbb{R}$, then LOTUS extends to the multidimensional case:

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy$$

The joint PDF must satisfy the following:

$$P(a \le X \le b, c \le Y \le d) = \int_a^b\int_c^d f_{X,Y}(x,y)\,dy\,dx$$

Group 12: Conditional Distributions, Conditional Expectation, Covariance and Correlation

1 Let X, Y be discrete random variables.
The **conditional PMF** of X given Y is:

$$p_{X|Y}(x|y) = P(X = x|Y = y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{p_{Y|X}(y|x)p_X(x)}{p_Y(y)}$$

2 Let X, Y be continuous random variables.
The **conditional PDF** of X given Y is:

$$f_{X|Y}(u|v) = \frac{f_{X,Y}(u,v)}{f_Y(v)} = \frac{f_{Y|X}(v|u)f_X(u)}{f_Y(v)}$$

Joint / Marginal

3 Let X, Y be jointly distributed random variables.
If X is discrete, then we define the **conditional expectation** of $g(X)$ given $Y = y$ as:

$$E[g(X)|Y = y] = \sum_{x \in \Omega_X} g(x)p_{X|Y}(x,y)$$

If X is continuous, then we define the **conditional expectation** of $g(X)$ given $Y = y$ as:

$$E[g(X)|Y = y] = \int_{-\infty}^{\infty} g(x)f_{X|Y}(x|y)dx$$

4 Let X, Y be jointly distributed random variables.
If Y is discrete, then:

$$E[g(X)] = \sum_{y \in \Omega_Y} E[g(X)|Y = y]p_Y(y)$$

If Y is continuous, then:

$$E[g(X)] = \int_{-\infty}^{\infty} E[g(X)|Y = y]f_Y(y)dy$$

Expectation

Law of total Expectation:

Definition of covar

5 Let X, Y be random variables. The covariance of X and Y is:

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])].$$
$$= E[XY] - E[X]E[Y]$$

Properties of covar

6 Covariance satisfies the following properties:

1. If X, Y are independent then $Cov(X,Y) = 0$
   $Cov(X,X) = \mathrm{Var}(X)$
   $Cov(X,Y) = Cov(Y,X)$
   $Cov(aX + b, cY + d) = acCov(X,Y)$
   $Cov(X_1 + X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$
   $Var(X + Y) = Var(X) + Var(Y) + Cov(X,Y)$

Correlation

7

- Correlation is a way to remove the scale from the covariance:

$$Cor(X,Y) = \rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

- Properties:
  - $\rho(X,Y)$ is the covariance of standardizations of X and Y.
  - $-1 \leq \rho(X,Y) \leq 1$
  - $\rho(X,Y) = \pm 1$ if and only if $Y = aX + b$

$Var(Y) =$

$\rightarrow Y = \sum X_i$