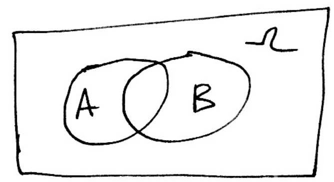
Stat 134 Review

<u>Equally likely outcomes</u>: For a set of equally likely outcomes, Ω, the probability of obtaining outcome A is,

$$\boxed{P(A) = \frac{\#(A)}{\#(\Omega)}}$$

<u>Events as sets</u>:



outcome space $\longrightarrow$ Ω

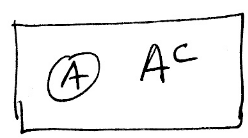
event (subset of Ω) $\longrightarrow$ A, B, C, etc.

not A (A complement) $\longrightarrow$ A^c

either A or B or both $\longrightarrow$ $A \cup B$

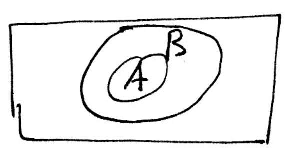
both A and B $\longrightarrow$ $A \cap B$ or AB

<u>Complements</u>: $\boxed{P(A^c) = 1 - P(A)}$



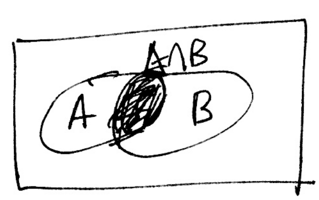
<u>Difference rule</u>:

If $A \subset B$, then $P(B) - P(A) = P(B \cap A^c)$



<u>Inclusion Exclusion Rule</u>:

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$



Generalized (multiple sets)

① Include each set

② Exclude unique pairwise intersections

③ Include unique triplet-wise intersections,

④ ...

Note: Demorgan's laws,

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

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

Conditional probability

$$\boxed{P(A \cap B) = P(B|A)\,P(A) = P(A|B)\,P(B)}$$ [multiplication rule]

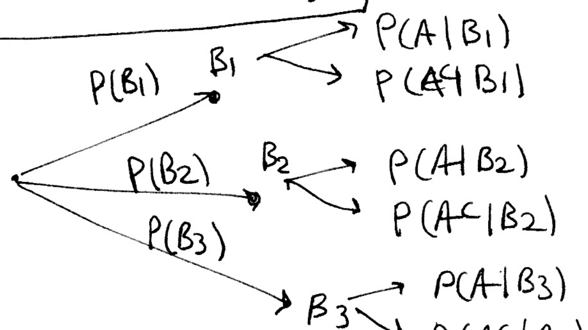
This leads us to Bayes rule,

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

Rule of Average conditional probabilites:

For a partition $B_1, \ldots, B_n$ of Ω,

$$\boxed{P(A) = \sum_{j=1}^{n} P(A|B_j)\,P(B_j)}$$



Generalizing Bayes Rule:

For a partition $B_1, \ldots, B_n$ of Ω, using our result above

$$P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(A|B_i)\,P(B_i)}{\sum_{j=1}^{n} P(A|B_j)\,P(B_j)}$$

Independence:

$$P(A|B) = A \iff P(B|A) = P(B) \iff P(A \cap B) = P(A)\,P(B)$$

For independence of n events, we must have,

$P(A_1 \cap \ldots \cap A_n) = P(A_1) \cdots P(A_n)$ for all possible intersections

For example: 3 events, A, B, and C are independent iff,

① A, B, C are pairwise independent

② $P(A \cap B \cap C) = P(A)\,P(B)\,P(C)$

Trials and Sampling

Bernoulli Trial: suppose you throw a coin with probability p of getting heads.

$$\text{Bernoulli}(p) = \begin{cases} 0, & q = 1-p \\ 1, & p \end{cases}$$

↑ value ↑ probability

Binomial Distribution:

For n independent Bernoulli(p) trials, the binomial dist. gives the probability of k successes.

$$\boxed{P(k \text{ success in } n \text{ trials}) = \binom{n}{k} p^k q^{n-k}}$$

number of ways to choose n successes in k trials

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

One way to think about the binomial dist is via the binomial expansion: $(p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$

Mean of binomial: $\mu = np$ [pf with indicators]

Variance of binomial: $\sigma^2 = npq$

Mode of binomial: $m = \lfloor np + p \rfloor$

Normal Approx. to Binomial

<u>Normal distribution</u>: the normal curve is given by

$$y = \frac{1}{\sqrt{2\pi}\,\underset{\uparrow\,SD}{\sigma}} e^{-\frac{1}{2}(x-\underset{\uparrow\,mean}{\mu})^2}$$

where $\mu = np$; $\sigma = \sqrt{npq}$

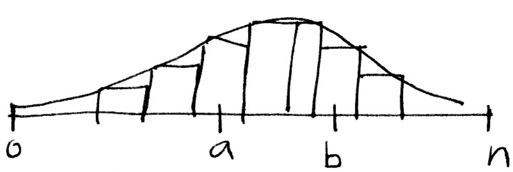
<u>Standard normal density</u>:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad \text{where } z = \frac{x-\mu}{\sigma}$$

<u>Cumulative dist. Function</u>:

$$\Phi(z) = \int_{-\infty}^{z} \phi(z')\,dz'$$

For n sufficiently large ~~and~~ (i.e. $n > 20$) and $\mu - 3\sigma > 0$, $\mu + 3\sigma < n$, then we can approximate the binomial distribution as normal.



$$P(a<k<b) = \Phi\left(\frac{(b+1/2)-\mu}{\sigma}\right) - \Phi\left(\frac{(a-1/2)-\mu}{\sigma}\right)$$

~~Binomial~~

Poisson Approx. to Binomial

For p very small (or p close to 1), even for large n, the binomial dist. is no longer symmetric $\Longrightarrow$ normal approx no longer holds.

$P = 1/8$ $\qquad$ $P = 1/2$ $\qquad$ $P = 7/8$

$n = 8$

$\approx$ Poisson $\qquad$ $\approx$ Normal $\qquad$ $\approx$ Poisson

In this regime, the poisson approx. the binomial. The probability of k successes, $P(k)$ becomes

$$P(k) = \frac{e^{-\mu}\mu^k}{k!} \quad \text{where } \mu = np, \; k = 0, 1, 2, \ldots$$

Random Sampling

Independent Trials (with replacement) → Binomial (2 outcomes per trial)

Independent Trials (with replacement) → Multinomial (many outcomes per trial)

Dependent Trials (without replacement) → Hypergeometric (2 outcomes)

Dependent Trials (without replacement) → multivariate Hypergeometric (many outcomes)

Suppose we have a population of N partitioned into categories $N_1, N_2, N_3, \ldots$ If we take n sample and want to find the probability of $n_1, n_2, n_3, \ldots$

<u>Multinomial</u>: (with replacement)

$$P(n_1, n_2, \ldots) = \binom{n}{n_1, n_2, \ldots} p_1^{n_1} p_2^{n_2} \cdots = \frac{n!}{n_1! \, n_2! \cdots} \left(\frac{N_1}{N}\right)^{n_1} \left(\frac{N_2}{N}\right)^{n_2} \cdots$$

<u>Multivariate Hypergeometric</u>: (without replacement)

Define $(N)_n = (N)(N-1)\cdots(N-n+1) = \binom{N}{n} n!$

$$P(n_1, n_2, \ldots) = \binom{n}{n_1, n_2, \ldots} \frac{(N_1)_{n_1} (N_2)_{n_2} \cdots}{(N)_n} = \frac{\binom{N_1}{n_1}\binom{N_2}{n_2}\cdots}{\binom{N}{n}}$$

Random Variables

A random variable is a stoichiastic quantity, where there is a defined probability for the RV to have some value.

Sum of independent Poisson is Poisson :

Suppose RVs X and Y are defined,

$X \sim \text{Poisson}(\mu = \mu_1)$; $Y \sim \text{Poisson}(\mu = \mu_2)$; $S = X + Y$

$$P(S=s) = \sum_{k=0}^{s} P(X=k)P(Y=s-k)$$

$$= \sum_{k=0}^{s} \frac{e^{-\mu_1}\mu_1^k}{k!} \frac{e^{-\mu_2}\mu_2^{(s-k)}}{(s-k)!} = \sum_{k=0}^{s} e^{-(\mu_1+\mu_2)} \frac{\mu_1^k \mu_2^{(s-k)}}{k!\,(s-k)!}$$

$$= e^{-(\mu_1+\mu_2)} \frac{1}{s!} \underbrace{\sum_{k=0}^{s} \frac{s!}{k!\,(s-k)!} \mu_1^k \mu_2^{(s-k)}}_{(\mu_1+\mu_2)^s} = \frac{e^{-(\mu_1+\mu_2)}(\mu_1+\mu_2)^s}{s!}$$

Thus, $S \sim \text{Poisson}(\mu_1+\mu_2)$

$$\boxed{\text{Poisson}(\mu_1) + \text{Poisson}(\mu_2) = \text{Poisson}(\mu_1+\mu_2)}$$ [assuming independence]

Expectation

$$E(X) = \sum_{x'} x' P(X=x')$$

① $E(c) = c$ for c constant

② $E(X+Y) = E(X) + E(Y)$ [X, Y need not be independent]

③ $E(aX+b) = aE(X) + b$

Pf) Addition Rule:

$$E(X+Y) = \sum_{x',y'} (x'+y') P(X=x' \wedge Y=y')$$

$$= \left(\sum_{x',y'} x' P(X=x' \wedge Y=y')\right) + \left(\sum_{x',y'} y' P(X=x' \wedge Y=y')\right)$$

$$= \sum_{x'} x' \left(\sum_{y'} P(X=x' \wedge Y=y')\right) + \sum_{y'} y' \left(\sum_{x'} P(X=x' \wedge Y=y')\right)$$

$$= \sum_{x'} x' P(X=x') + \sum_{y'} y' P(Y=y') = E(X) + E(Y) \quad \square$$

Some further properties of expectation:

* $E(g(X)) = \sum_{x'} g(x') P(X=x') \neq g(E(X))$ in general

* In general, $E(X \cdot Y) \neq E(X) \cdot E(Y)$. If

$$E(X \cdot Y) = E(X) \cdot E(Y) \Longleftrightarrow X, Y \text{ are independent}$$

Method of Indicators:

An indicator for an event ~~[illegible]~~ is a random variable with value 1 with probability the event occurs and value 0 with probability that the event does not occur,

$$I_j = \begin{cases} 1, & P(j) \text{ that event } j \text{ occurs} \\ 0, & \text{else} \end{cases}$$

$$E(I_j) = 1 \cdot P(j) + 0 \cdot P(\text{not } j) = P(j)$$

We can calculate expectation of a RV by breaking it into a sum of indicators,

Example: ~~A~~ 12 people get into an elevator. Each person will exit at one of 10 floors chosen randomly and independently. What is the expected number of floors the elevator will stop?

$X = \#$ floors the elevator stops

$$I_j = \begin{cases} 1, & \text{at least 1 person gets off at floor } j \\ 0, & \text{else} \end{cases}$$

$P(I_j) = P(\text{at least one person gets off at floor } j)$ is unconditioned on whether someone has gotten ~~up~~ off at a previous floor certainly the ~~events~~ RVs I_i, I_j are not independent, but that does not matter when calculating expectation.

$$E(X) = \cancel{E(I_1)} E(I_1 + I_2 + \dots I_{10}) = E(I_1) + E(I_2) + \dots E(I_{10})$$

$$= 10\, E(I_1) = 10 \left[1 - (9/10)^{12}\right]$$

Tail Sum Formula for Expectation:

For RV X with values $0, 1, \dots, n$

$$E(X) = 0 \cdot P(X=0) + 1 \cdot P(X=1) + 2P(X=2) + \dots + n \cdot P(X=n)$$

$$= P(X=1) + P(X=2) + \dots + P(X=n)$$
$$+ P(X=2) + \dots + P(X=n)$$
$$+ \dots$$

$$\boxed{E(X) = \sum_{j=1}^{n} P(X \geq j)}$$

Variance

Variance and SD: Variance is the expectation of the deviation squared.

$$\boxed{\begin{aligned} Var(X) &= E((X-\mu)^2) \\ SD(X) &= \sqrt{Var(X)} \end{aligned}}$$

Note: ① $Var(aX+b) = a^2 Var(X)$

② If X, Y independent,
$Var(X+Y) = Var(X) + Var(Y)$

We can also express variance as:

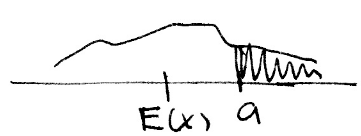
$$E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2] = E(X^2) - 2\mu E(X) + \mu^2$$
$$= E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2$$

$$\boxed{Var(X) = E(X^2) - [E(X)]^2}$$

Inequalities

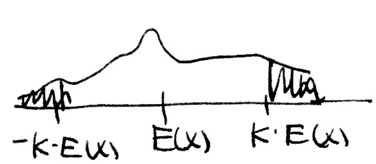
① Markov's inequality: If $X \geq 0$, then,

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



② Chebychev's inequality,

$$P\left(|X - E(X)| \geq k \cdot SD(X)\right) \leq \frac{1}{k^2}$$



Central limit Theorem

let $S_n = X_1 + X_2 + \dots + X_n$ be the sum of independent, identical distributions. Then, for large n, S_n is approx. normal with $\mu = n E(X)$ and $\sigma = \sqrt{n}\, SD(X)$

Poisson Distribution

The poisson dist models the number of occurrences of some event with very low probability.

$$\binom{n}{k} p^k (1-p)^{n-k} \longrightarrow \frac{e^{-\mu}\mu^k}{k!} \quad \text{as } n \to \infty,\ p \to 0 \text{ with } \mu = np$$

Poisson mean: $E(X) = \mu$

Poisson variance: $Var(X) = \mu$

Sum of indep poisson is poisson: see page 6.

Random Scatter

Suppose raindrops are hitting a square. If we divide the square into many smaller squares, and,

① There can only be 1 hit max at one subdivision

② Each subdivision is independent and we know

$$\lambda = \text{average \#hits/Area} = \text{intensity}$$

Then each subdivision can be modeled as a poisson process (the chance of a drop hitting any one subdivision is small).

The sum of iid poisson is poisson, so for $X = \#\text{hits}/\text{Area}$

$$X \sim \text{poisson}(\lambda A)$$

Poisson thinning:

Suppose that in a poisson scatter with intensity λ, each point of the scatter is kept with probability p and thinned with probability $1-p$. Then, for,

$$X = \#\text{ scatter points kept},$$
$$X \sim \text{Poisson}(\lambda p)$$

Continuous Distributions

*Notation: $f_X(x)$ = density of RV X
$F_X(x)$ = cdf of RV X

Probability density: For a continuous RV, the probability of the random value having some range of values is given by the probability density $f(x)$

$$P(a < x < b) = \int_a^b f(x)\,dx$$

$$P(X \in dx) = f(x)\,dx$$

probability per unit length ~~that~~ for values near x.

Exponential Distribution

Models the time of a ~~stoichiastic~~ stoichiastic process that experiences no ageing effect. If T is a random time with rate λ,

$$T \sim \text{exponential}(\lambda)$$
$$f(t) = \lambda e^{-\lambda t} \quad ; t \geq 0$$
$$E(T) = SD(T) = 1/\lambda$$

Exponential survival function: $P(T > t) = e^{-\lambda t}$

This implies that T has a cdf of,

$$F(t) = 1 - e^{-\lambda t}$$

Memoryless property of Exp(λ):

Given a survival time t, the chance of surviving a further time s is equal to surviving to time s from the start. We say "survival time" because we imagine T = time until lightbulb fails.

$$P(T>t+s \mid T>t) = \frac{P(T>t+s \cap T>t)}{P(T>t)}$$

$$= \frac{P(T>t+s)}{P(T>t)} = \frac{e^{-(t+s)\lambda}}{e^{-t\lambda}} = e^{-s\lambda}$$

$$\boxed{P(T>t+s \mid T>t) = P(T>t)}$$

Relating exp(λ) to poisson(λ):

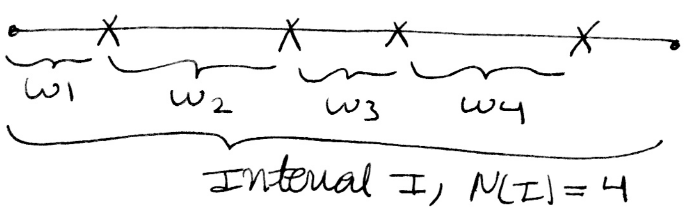
Suppose we count the number of arrivals in a fixed interval of time I. The number of arrivals in the interval is,

$$N(I) \sim \text{Poisson}(\lambda I)$$

The time between ~~intervals~~ arrivals are independent

W_i = time between i and $i+1$ arrival

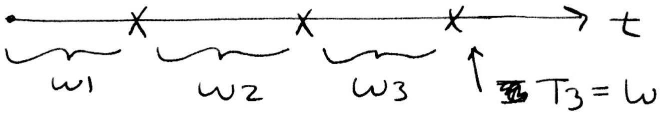
$$W_i \sim \exp(\lambda)$$



Gamma Distribution

Define T_r = the time of the r^{th} arrival of a poisson process

$$T_r = W_1 + W_2 + \dots + W_r \quad \text{where } W_j = \text{iid } \exp(\lambda)$$



$T_3 = W_1 + W_2 + W_3$

To calculate the density lets also define

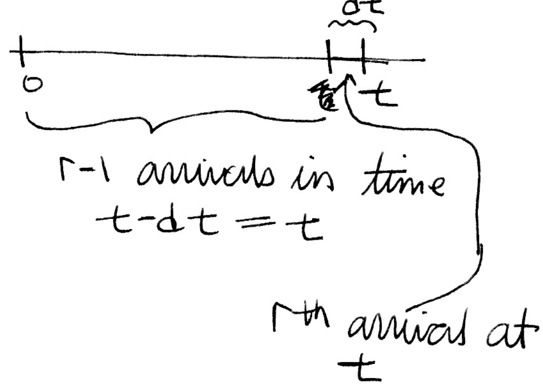
N_t = # arrivals in time t

$N_t \sim \exp(\lambda)$

Thus,

$$P(T_r \in dt) = P(N_t = r-1)\, P(T_1 = 0)$$

$$P(T_r \in dt) = \left(\frac{e^{-\lambda t}(\lambda t)^{r-1}}{(r-1)!}\right)\left(\lambda e^{0}\, dt\right)$$



$$\boxed{f(t) = \frac{e^{-\lambda t}(\lambda t)^{r-1}}{(r-1)!}\lambda}$$

gamma (r, λ)

time of r^{th} arrival; rate

Right-tail Probability:

$$P(T_r > t) = P(N_t \le r-1) = \sum_{k=0}^{r-1} \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

Mean and SD of gamma(r, λ):

$$E(T_r) = r/\lambda \;; \quad SD(T_r) = \frac{\sqrt{r}}{\lambda}$$

Generalize to non-integer r:

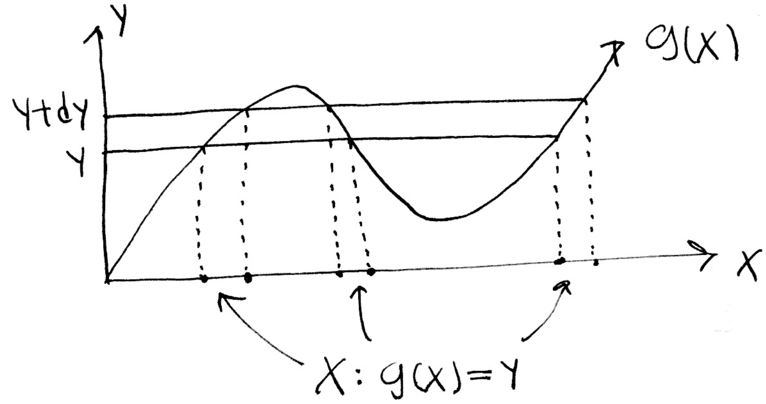
We generalize by keeping the variable form and introducing a normalizing factor.

$$\boxed{f_{r,\lambda}(t) = \frac{\lambda^r t^{r-1} e^{-\lambda t}}{\Gamma(r)}} \quad \text{for } t \ge 0$$

$$\boxed{\Gamma(r) = \int_0^\infty t^{r-1} e^{-t}\, dt} \;; \quad \Gamma(r) = (r-1)! \text{ for } r \in \mathbb{N}$$

Change of Variable

Suppose we have RV, X, with probability density $f_X(x)$.
If we define $Y = g(X)$, then what is $f_Y(Y)$?



Thus, $$f_Y(Y)\,dY = \sum_{\{x:g(x)=y\}} f_X(x)\,dx$$

$$\boxed{f_Y(Y) = \sum_{X:g(x)=Y} \frac{f_X(X)}{|dY/dX|}}$$

$\uparrow \left|\frac{dY}{dX}\right| = \left|\frac{d}{dx} g(x)\right|$ (abs. value because the derivative can be negative)

<u>In the linear case</u>:

$$f_Y = f_{ax+b}(Y) = \frac{1}{|a|} f_X\left(\frac{Y-b}{a}\right)$$

[note $f_{ax+b}(Y)$ means the Prob density of the RV $ax+b$ as a function of Y]

Moment Generating Function

The moment generating function (MGF) is defined,

$$\boxed{M_X(t) = E(e^{Xt})}$$

$$E(e^{Xt}) = E\left(1 + Xt + \tfrac{1}{2}(Xt)^2 + \dots\right) = E(1) + E(Xt) + \tfrac{1}{2}E((Xt)^2) + \dots$$

$$M_X(t) = E(1) + tE(X) + \tfrac{1}{2}t^2E(X^2) + \dots$$

Thus, $$\boxed{E(X^k) = \frac{d^k}{dt^k}(M_X(t))\Big|_{t=0}}$$

Properties of MGF:

① If X and Y are independent, $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$

② MGF uniquely specified a distribution,

$$M_X(t) = M_Y(t) \iff f_X = f_Y$$

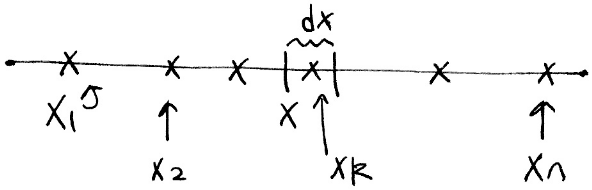
Order Statistics

let $X_1, X_2, \ldots, X_n$ be random variables that are independent, identically distributed. We can relabel the x's so that

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$$

$X_{(k)}$ is called the kth <u>order statistic</u>

<u>Density of the kth order statistic</u>:



$$\boxed{f_{(k)}(x)dx = \binom{n}{k-1,\,1,\,n-k} F(x)^{k-1} f(x)dx\,(1-F(x))^{n-k}}$$

Beta Distribution

The kth order ~~distribution~~ statistic of n independent Uniform(0,1) random variables is defined to have distribution beta$(k, n-k+1)$

$$f_{U(k)} = \binom{n}{k-1,\,1,\,(n-k+1)-1} x^{k-1}(1-x)^{(n-k+1)-1}$$

Generalizing,

$$\boxed{\begin{array}{l}\dfrac{1}{B(r,s)} x^{r-1}(1-x)^{s-1} \quad \text{for } 0<x<1 \\ B(r,s) = \int_0^1 x^{r-1}(1-x)^{s-1}dx\end{array}} \quad [\text{beta}(r,s)]$$

For, $r,s \in \mathbb{N}$, $B(r,s) = \dfrac{(r-1)!(s-1)!}{(r+s-1)!} = \dfrac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}$

Rayleigh Distribution

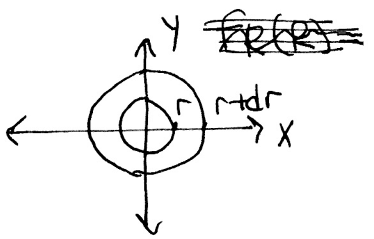
Suppose X and Y are iid standard normal: each has normal density,

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

Thus, X and Y have joint density,

$$f(x,y) = \phi(x)\phi(y) = \frac{1}{2\pi} e^{-1/2(x^2+y^2)}$$

Suppose we define a new RV, $R = \sqrt{X^2+Y^2}$. The probability density of R is given by,



~~FR(R)~~

$$P(R \in dr) = (2\pi r\, dr) \frac{1}{2\pi} e^{-1/2(x^2+y^2)}$$
$$= r e^{-1/2 r^2} dr$$

$$\boxed{f_R(r) = r e^{-1/2 r^2}}$$ [Rayleigh dist]

where $R = \sqrt{X^2+Y^2}$ and

$X, Y \sim$ iid standard normal

<u>Cdf of Rayleigh Dist</u>: $F_R(r) = \int_0^r r' e^{-1/2 r'^2} dr' = 1 - e^{-1/2 r^2}$

$$\boxed{F_R(r) = 1 - e^{-1/2 r^2}}$$

<u>Chi-square dist.</u>:

Define $R_n = \sqrt{Z_1^2 + \dots + Z_n^2}$

We can think of this as an n-dimensional sphere.

Thus,
$$P(R_n \in dr) = c_n r^{n-1} \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-1/2 r^2} dr$$

$c_2 = 2\pi$

$c_3 = 4\pi$

$\vdots$

Operations

Suppose we have random variables X and Y, and another random variable is defined as a function of X, Y.

E.g. $X \cdot Y$, X/Y, $\min(X,Y)$. What is the prob. density of Z?

<u>Method of cdf</u>:

① Determine $P(Z \leq z) = F_Z(z)$

② Then, $f_Z(z) = \frac{\partial}{\partial z}(F_Z(z))$

<u>Density of X+Y</u>:

If (X,Y) has joint density $f(x,y)$, and $z = x+y$,

$$\boxed{f_{X+Y}(z) = \int_{-\infty}^{\infty} f(x, z-x)\,dx}$$

If X and Y are independent,

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

<u>Density of Ratios</u>:

Let $Z = Y/X$.

$$P(X \in dx \cap Y/X \in dz) = f(x, xz)\,|x|\,dx\,dz$$

Jacobian

$$\boxed{f_{Y/X}(z) = \int_0^{\infty} |x| f(x, xz)\,dx}$$

Conditional Distributions

Conditional dist of Y given $X = x$:

$$P(Y=y) = \sum_{x} \underbrace{P(Y=y \mid X=x)\, P(X=x)}_{P(X=x \,\cap\, Y=y)}$$

In the continuous case, the prob of some event A,

$$P(A) = \int P(A \mid X=x) f_X(x)\, dx$$

Conditional Expectation:

The conditional expectation of a RV Y given an event A,

$$\boxed{E(Y \mid A) = \sum_{\text{all } y} y\, P(Y=y \mid A)}$$

Properties:

① $E(X+Y \mid A) = E(X \mid A) + E(Y \mid A)$

② $E(Y) = \sum_{j=1}^{n} E(Y \mid A_j) P(A_j)$

Rule of Average Conditional Expectation:

$$\boxed{\begin{aligned} E(Y) &= \sum_{\text{all } x} E(Y \mid X=x) P(X=x) \\ E(Y) &= E[E(Y \mid X)] \end{aligned}}$$

Covariance

The covariance is defined,

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

Alternatively, $\mathrm{Cov}(X,Y) = E[XY - \mu_X Y - X\mu_Y + \mu_X\mu_Y] = E(XY) - \mu_X\mu_Y - \mu_X\mu_Y + \mu_X\mu_Y$

$$\boxed{\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)}$$

Additionally,

$$\mathrm{Var}(X+Y) = E[[(X-\mu_X)+(Y-\mu_Y)]^2]$$

$$= E[(X-\mu_X)^2] + E[(Y-\mu_Y)^2] + 2E[(X-\mu_X)(Y-\mu_Y)]$$

$$\boxed{\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)}$$

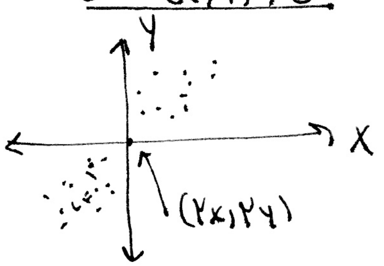
Properties of covariance:

1. $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$
2. $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$
3. $\mathrm{Cov}(X,c) = 0$ ← const.
4. $\mathrm{Cov}(X+Y,Z) = \mathrm{Cov}(X,Z) + \mathrm{Cov}(Y,Z)$
5. $\mathrm{Cov}(aX,bY) = ab\,\mathrm{Cov}(X,Y)$

4. If X and Y are independent,

$\mathrm{Cov}(X,Y) = 0$

$\mathrm{Cov}(X,Y) > 0$



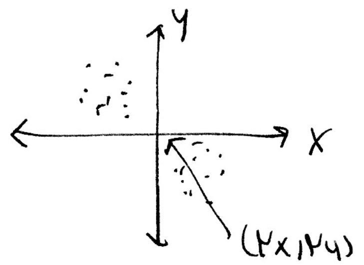
above average X is associated with above average Y; below X with below Y

$\mathrm{Cov}(X,Y) = 0$

X, Y are independent

no association between X and Y

$\mathrm{Cov}(X,Y) < 0$



above average X is associated with below average Y, and vice versa.

Correlation

$$\boxed{\mathrm{Corr}(X,Y) = E\left[\left(\frac{X-\mu_X}{SD_X}\right)\left(\frac{Y-\mu_Y}{SD_Y}\right)\right] = \frac{\mathrm{Cov}(X,Y)}{SD(X)\,SD(Y)}}$$

$-1 \le \mathrm{Corr}(X,Y) \le 1$

Standard coordinates

Bivariate Normal