By: Gan Chin Yao l 2019 May

Joint probabilities

$f_{X,Y}(x,y) = P(X=x, Y=y)$ Discrete

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

Properties of Joint probability mass function

1. $f_{X,Y}(x,y) \ge 0$, for all $(x,y) \in R_{X,Y}$.
2. $\sum_x \sum_y f_{X,Y}(x,y) = \sum_x \sum_y P(X=x, Y=y) = 1.$

Let A be any set consisting of pairs of (x,y) values. Then the probability $P((X,Y) \in A))$ is defined by summing the joint probability mass function over pairs in A:

$$P((X,Y) \in A) = \sum_{(x,y) \in A} f_{X,Y}(x,y).$$

(b) For $f_X(x) > 0$, $f_{X,Y}(x,y) = f_{Y|X}(y|x) f_X(x)$.
For $f_Y(y) > 0$, $f_{X,Y}(x,y) = f_{X|Y}(x|y) f_Y(y)$.

Properties of Joint probability distribution function

1. $f_{X,Y}(x,y) \ge 0$, for all $(x,y) \in R_{X,Y}$.
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy\,dx = 1.$

Marginal Probability Mass

$$f_X(x) = P(X=x) = \sum_y P(X=x, Y=y) = \sum_y f_{X,Y}(x,y).$$

x is fix, y is changing

$$f_Y(y) = P(Y=y) = \sum_x P(X=x, Y=y) = \sum_x f_{X,Y}(x,y).$$

y is fix, x is changing

$f_{X,Y}(x,y)$

y \ x	0	1	2	3	4	5	$f_Y(y)$
0	0	0.01	0.02	0.05	0.06	0.08	**0.22**
1	0.01	0.03	0.04	0.05	0.05	0.07	**0.25**
2	0.02	0.03	0.05	0.06	0.06	0.07	**0.29**
3	0.02	0.04	0.03	0.04	0.06	0.05	**0.24**
$f_X(x)$	**0.05**	**0.11**	**0.14**	**0.20**	**0.23**	**0.27**	1

Marginal Probability Continuous

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

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

Conditional Probability Mass/density

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \text{ provided } f_Y(y) > 0,$$

given Y = y , for each y within the range of Y

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \text{ provided } f_X(x) > 0,$$

given X = x , for each x within the range of X.

e.g

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)\,dy = \begin{cases} \int_0^2 1/2\,dy, & \text{for } 0 \le x \le 2 \\ 0, & \text{otherwise} \end{cases}$$

Independent

Random variables X and Y are **independent** if and only if

$f_{X,Y}(x,y) = f_X(x) f_Y(y)$ for all x and y.

or equivalently $f_{X|Y}(x|y) = f_X(x)$ similar for f(x1, x2, ... xn)

Expectation of g(X, Y)

In the discrete case,

$$E[g(X,Y)] = \sum_x \sum_y g(x,y) f_{X,Y}(x,y).$$

In the continuous case,

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

Covariance

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

Properties of Covariance

$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = E(XY) - \mu_X \mu_Y.$
$\mathrm{Cov}(X,X) = V(X).$
$\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X).$
$\mathrm{Cov}(aX+b, cY+d) = ac\,\mathrm{Cov}(X,Y).$
$V(aX+bY) = a^2 V(X) + b^2 V(Y) + 2ab\,\mathrm{Cov}(X,Y)$
If X, Y are independent, then Cov(X,Y) = 0.
$\mathrm{cov}(aX, bY) = ab\,\mathrm{cov}(X,Y)$
$\mathrm{cov}(X+a, Y+b) = \mathrm{cov}(X,Y)$
$\mathrm{cov}(aX+bY, cW+dV) = ac\,\mathrm{cov}(X,W) + ad\,\mathrm{cov}(X,V) + bc\,\mathrm{cov}(Y,W) + bd\,\mathrm{cov}(Y,V)$

Correlation Coefficient

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}.$$

$-1 \le \rho_{X,Y} \le 1.$

$\rho_{X,Y}$ is a measure of the degree of linear relationship between X and Y.
If X and Y are independent, then $\rho_{X,Y} = 0$.
On the other hand, $\rho_{X,Y} = 0$ does not imply independence

Examples

$$E(XY) = \int_0^2 \int_0^1 xy\left(x^2 + \frac{xy}{3}\right)dx\,dy$$

find marginal f(x) first

$$E(X) = \int_0^1 x\left(2x^2 + \frac{2x}{3}\right)dx$$

$$E(X) = \int_0^1 x\left(2x^2 + \frac{2x}{3}\right)dx$$

P(Y | X = 0.5)

$$E(Y|X = 1/2) = \int_0^2 y\left(\frac{3+2y}{10}\right)dy$$

Discrete Uniform Distribution

If the random variable X assumes the values x1,x2,...,xk with equal probability, X is said to have a discrete uniform distribution and pmf:

$$f_X(x) = P(X=x) = \begin{cases} 1/k, & x = x_1, x_2, \ldots, x_k \\ 0, & \text{otherwise} \end{cases}.$$

Mean and variance:

$$\mu = E(X) = \sum x f_X(x) = \frac{1}{k}\sum_{i=1}^k x_i,$$

$$\sigma^2 = V(X) = \sum (x-\mu)^2 f_X(x) = \frac{1}{k}\sum_{i=1}^k (x_i - \mu)^2, \text{ or}$$

$$\sigma^2 = E(X^2) - \mu^2 = \frac{1}{k}\sum_{i=1}^k x_i^2 - \mu^2.$$

Bernoulli Distribution

A random variable X is defined to have a Bernoulli distribution with parameter 0 < p < 1, when it has probability mass function given as

$$f_X(x) = P(X=x) = p^x(1-p)^{1-x}, \quad \text{for } x = 0, 1.$$

E(X) = p and V(X) = p(1− p)

Binomial Distribution (Disrete)

Note: binomial distribution has finite number of successful outcome

Definition 4.9 (Binomial distribution) discrete

A random variable X is defined to have a **binomial distribution** with parameters $n \in \mathbb{Z}^+$ and $0 < p < 1$, written as $X \sim B(n,p)$, when it has probability mass function given as

n= number of independent trials
p = probability of a successful outcome

$$f_X(x) = P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad \text{for } x = 0, 1, 2, \ldots, n.$$

It can be shown that $E(X) = np$, and $V(X) = np(1-p)$.

I. independent
O. only 2 outcomes
C. probability of success is constant

X: number of successes in a sequence of n independent experiments, with p success

Geometric Distribution (Discrete)

P(X >n)= q^n Discrete, memoryless, try until you succeed

Definition 4.17 (Geometric distribution)

A random variable X is defined to have a **geometric distribution** with parameter $0 < p < 1$, written as $X \sim \mathrm{Geom}(p)$, when it has probability mass function given as

X: How many independent trials are needed until first succeed

$$f_X(x) = P(X=x) = (1-p)^{x-1}p, \quad \text{for } x = 1, 2, 3, \ldots.$$

It can be shown that $E(X) = \frac{1}{p}$ and $V(X) = \frac{1-p}{p^2}$. CDF: 1 - (1 - p)^k

i.e fail for x-1 times, then pass on last trial. Therefore, p = (q^x-1)p

Alternative formulation for the Geometric

Some books define the geometric random variable Y as the **number of failures encountered until the first success is achieved**. Thus

X: How many failures happen before first succeed

$$Y = X - 1.$$

and Y takes values $0, 1, 2, \ldots$. It follows that

$$P(Y = y) = (1-p)^y p, \quad \text{for } y = 0, 1, 2, \ldots,$$

and

CDF: 1 - (1-p)^(k+1)

$$E(Y) = \frac{1-p}{p}, \quad V(Y) = \frac{1-p}{p^2}.$$

Memoryless: P(X >n+k | X >n) = P(X >k), for all n, k ≥ 1.

Negatiave Binomial Distribution

number of trials you need in order to collect a fix number of successe

Definition 4.25 (Negative Binomial distribution)

A random variable X is defined to have a **Negative Binomial distribution** with parameters $k \in \mathbb{Z}^+$ and $0 < p < 1$, written as $X \sim \mathrm{NB}(k,p)$, when it has probability mass function given as

k = number of successes you want, p is probability for success

$$f_X(x) = P(X=x) = \binom{x-1}{k-1} p^k q^{x-k}, \quad \text{for } x = k, k+1, k+2, \ldots.$$

The probability that the kth success come from the x trial

It can be shown that $E(X) = \frac{k}{p}$ and $V(X) = \frac{(1-p)k}{p^2}$. Geom(p) = NB(1, p).

Your x-1 trials has k-1 success, and last one, i.e. x-th, is success

Poisson (Discrete)

Poisson Experiment Poisson is discrete

Experiments yielding numerical values of a random variable X, the number of successes occurring (i) during a given time interval or (ii) in a specified region, are called **Poisson experiments**.

R. random
I. independent
M. mean is constant

$$f_X(x) = P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad \text{for } x = 0, 1, 2, \ldots.$$

E (X) = λ , and V (X) = λ

Note: sum(P(X=r)) from r=0 to inifinite is 1

Only for independent poisson,
X ~ Po(m), Y ~ Po(n) => X + Y ~ Po(m + n)
Can only sum, cannot - * / Also applies for more than 2

~approx BINOMIAL to POISSON

Let X be a Binomial random variable with parameters n and p.
Suppose that n → ∞ and p→ 0 in such a way that λ = np remains a constant as n→∞
Then X will have approximate a Poisson distribution with parameter np

$$\lim_{\substack{n\to\infty \\ p\to 0}} P(X=x) = \frac{e^{-np}(np)^x}{x!}.$$

The approximation is good when n ≥ 20 and p≤0.05, or if n ≥ 100 and np ≤ 10. i.e. X~B(n, p) ---> X~Po(np)

Continuous Uniform Distribution

A random variable X is said to follow a **uniform distribution** over the interval $[a,b]$, denoted by $X \sim U(a,b)$, if its probability density function is given by

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{for } a \le x \le b \\ 0, & \text{otherwise} \end{cases}.$$

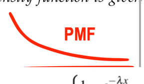
It can be shown that $E(X) = \frac{a+b}{2}$ and $\mathrm{var}(X) = \frac{(b-a)^2}{12}$.

Exponential Distribution

P(X>s+t|X>s)=P(X>t), fors,t>0.

A random variable X is said to follow an **exponential distribution** with parameter $\lambda > 0$, denoted by $X \sim \mathrm{Exp}(\lambda)$, if its probability density function is given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{for } x > 0 \\ 0, & \text{for } x \le 0 \end{cases}$$



It can be shown that $E(X) = \frac{1}{\lambda}$ and $\mathrm{var}(X) = \frac{1}{\lambda^2}$. CDF $F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \le 0 \end{cases}.$

Normal Distribution

A random variable X is said to follow a **normal distribution** with parameters $-\infty < \mu < \infty$ and $\sigma > 0$, denoted by $X \sim N(\mu, \sigma^2)$, if its probability density function is given by

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

It can be shown that $E(X) = \mu$ and $V(X) = \sigma^2$.

Standard normal:

pdf $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$; **cdf** $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x e^{-y^2/2}\,dy.$ $\boxed{\frac{b-\mu}{\sigma}}$

IfY~N(μ,σ2),then X = (Y−μ)/σ ~ N(0,1) If X ~N(0,1),thenY =aX+b~N(b,a²)for a,b belong R.

Quantile

qth quantile of the random variable X is the number z_q that satisfies P(X ≤ z_q)=q.
Table: means P(Z < 1.960) = 0.975

~approx. Binomial to Normal

Continuity Correction

Suppose X is a binomial random variable with mean $\mu = np$ and variance $\sigma^2 = np(1-p)$. Then as $n \to \infty$, $Z = \frac{X - np}{\sqrt{np(1-p)}}$ is approximately distributed as $N(0,1)$.

When n ⟶ ∞ and p ⟶ 0.5
np>5 and n(1-p)>5.

Law of Large Number

If $X_1, \ldots, X_n$ be a random sample of size n from a population with mean μ and finite variance σ^2. Then for any $\varepsilon \in \mathbb{R}$, $P(|\overline{X} - \mu| > \varepsilon) \to 0$ as $n \to \infty$.

Central Limit Theorem (CLT)

n ≥ 30

Let $X_1, \ldots, X_n$ be a random sample from a population with mean μ and finite variance σ^2. The sampling distribution of the sample mean $\overline{X}$ is approximately normal with mean μ and variance σ^2/n if n is sufficiently large.

This means that

mean only, not variance, not min, not max

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \text{ approximately,}$$

crucial is sample size large, not number of repetition of small sample size

or equivalently

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ approximately.}$$

TO NORMAL

If Xi, i = 1,2,...,n are N(μ,σ²), then Xbar is N(μ,σ² /n) regardless of the sample size n.
Similar as ^, if original was approx normal, then Xbar also approx normal

Difference of 2 means

Xn >= 30, Yn >= 30, X&Y indepndent

$$E(\overline{X} - \overline{Y}) = \mu_{\overline{X}-\overline{Y}} = \mu_1 - \mu_2 \quad \text{and} \quad V(\overline{X} - \overline{Y}) = \sigma^2_{\overline{X}-\overline{Y}} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

$$Z = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1) \text{ approximately.}$$

Gamma Function

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y}\,dy.$$

It can be shown,via integration by parts, that Γ(α)=(α−1)Γ(α−1).

$\Gamma(1) = \int_0^\infty e^{-y}\,dy = 1.$ For integral values of α, say, $\alpha = n = 1, 2, 3, \ldots, \Gamma(n) = (n-1)!.$

χ2-DISTRIBUTION

$$f_Y(y) = \frac{1}{2^{n/2}\Gamma(n/2)} y^{n/2-1} e^{-y/2}, \quad \text{for } y > 0.$$

n degrees of freedom

If Y~χ2(n), then E(Y)=n and V(Y)=2n   For large n, χ 2 (n) ~ N (n, 2n) approximately.
Y1~χ2(n1) + Y2~χ2(n2) + ... = Y~χ2(n1 + n2 + ...)
- if Z~N(0,1), then Z^2~χ2(1)

$$Y = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(n).$$

Xi is normal

Table: P(χ2(15) > 10.307) = 0.8; χ2(15,0.8) = 10.307

Sample Variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2.$$

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{\sigma^2}$$ ~ χ2(n - 1) <= any n, original must be normal

t-Distribution

Estimate pop param(e.g.mean) of normal distr, n is small, sigma(pop) is unknown (not variance)

Let Z be a standard normal variable and U a χ^2 random variable with n degrees of freedom. If Z and U are independent, then T given by $T = \frac{Z}{\sqrt{U/n}}$

$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

is a t-distribution with n degrees of freedom $, -\infty < t < \infty.$

Table: t(1-alpha) = -t(alpha) | t(alpha) = -t(1 - alpha) | df is sample size - 1
- If T~t(n), then E(T) = 0 and V(T) = n/(n-2) for n > 2

If the Xi's are normal, then

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$ ~ t(n - 1)

S = sample SD
(i.e sqroot sample variance)
Xbar = Sample mean
mu = Population mean
n = sample size (for all n, mainly small)

F-Distribution

Let U and V be independent random variables having χ^2 distributions with n_1 and n_2 degrees of freedom, respectively. The distribution of the random variable,

$$F = \frac{U/n_1}{V/n_2},$$

is called a F distribution with (n_1, n_2) degrees of freedom.

The probability density function of F is given as

$$f_F(x) = \frac{n_1^{n_1/2} n_2^{n_2/2}\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \frac{x^{n_1/2-1}}{(n_1 x + n_2)^{(n_1+n_2)/2}}, \quad \text{for } x > 0.$$

- If F ~ F(n,m), then 1/F ~ F(m,n)
- table: P(F > F(n1,n2;α)) = α, when F ~ F(n1,n2)
- F(5,4;0.05) = 6.26 means that P(F > 6.26) = 0.05, where F ~ F(5, 4)
- F(n1, n2; α) = 1/F(n2, n1; 1−α)

$$F = \frac{U/(n_1-1)}{V/(n_2-1)} = \frac{(n_1-1)S_1^2/(n_1-1)\sigma_1^2}{(n_2-1)S_2^2/(n_2-1)\sigma_2^2} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1).$$

Confidence

Estimator: Xbar, Estimate: actual xbar value

the interval $\hat\theta_L < \theta < \hat\theta_U$, computed from the selected sample is called a $(1-\alpha)100\%$ ***confidence interval*** *for* θ,
the fraction $(1-\alpha)$ *is called the* ***confidence coefficient*** *or* ***degree of confidence***, *and*
the end points $\hat\theta_L$ *and* $\hat\theta_U$ *are called the* ***lower and upper confidence limits*** *respectively.*

$$P(|\bar X - \mu| \le e) \ge 1-\alpha$$

Margin of error: e

e = |Xbar - mu|
n= sample size

$$n \ge \left(\frac{z_{\alpha/2}\cdot\sigma}{e}\right)^2$$

AP: (n/2)(a+L) = (n/2)(2a + (n-1)d)
GP = (a(1-r^n)) / (1-r)

Estimate	Know	Don't Know	Remarks	Formula	Workings
μ(pop)	σ2(pop) big n/normal		Zα/2	$\bar X \pm z_{\alpha/2}\frac{\sigma_{pop}}{\sqrt n}$	$P\left(-z_{\alpha/2} \le \frac{\bar X-\mu}{\sigma/\sqrt n} \le z_{\alpha/2}\right) = 1-\alpha$
μ(pop)	Pop normal small n	σ2(pop)	t~(n-1) s to estima	$\bar X \pm t_{n-1;\alpha/2}\frac{S}{\sqrt n}$	$P\left(-t_{n-1;\alpha/2} \le \frac{\bar X-\mu}{S/\sqrt n} \le t_{n-1;\alpha/2}\right) = 1-\alpha$
μ(pop)	Pop normal big n > 30	σ2(pop)	Zα/2 CLT s to estima	$\bar X \pm z_{\alpha/2}\frac{S}{\sqrt n}$	
μ1(pop) - μ2(pop)	σ_1^2, σ_2^2 σ_1^2 != σ_2^2 big n / normal		Zα/2 CLT	$(\bar X_1-\bar X_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma^2_{1pop}}{n_1}+\frac{\sigma^2_{2pop}}{n_2}}$	$P\left(-z_{\alpha/2} < \frac{(\bar X_1-\bar X_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} < z_{\alpha/2}\right) = 1-\alpha$
μ1(pop) - μ2(pop)	σ_1^2 != σ_2^2 Big n1, n2	σ_1^2, σ_2^2	Zα/2 CLT s to e	$(\bar X_1-\bar X_2) \pm z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$	$P\left(-t_{n_1+n_2-2;\alpha/2} < \frac{(\bar X_1-\bar X_2)-(\mu_1-\mu_2)}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} < t_{n_1+n_2-2;\alpha/2}\right) = 1-\alpha$
μ1(pop) - μ2(pop)	σ_1^2 == σ_2^2 Pop normal Small n1, n2	σ_1^2, σ_2^2	T-dist pooled S	$(\bar X_1-\bar X_2) \pm t_{n_1+n_2-2;\alpha/2}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$	$S_p^2 = \frac{\sum_{i=1}^{n_1}(X_i-\bar X)^2+\sum_{i=1}^{n_2}(Y_i-\bar Y)^2}{n_1+n_2-2} = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$
μ1(pop) - μ2(pop)	σ_1^2 == σ_2^2 Big n1, n2	σ_1^2, σ_2^2	Zα/2 CLT s to est	$(\bar X_1-\bar X_2) \pm z_{\alpha/2}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$	
μD = μ1(pop) - μ2(pop)	Pop normal Small n1, n2	σ_{Dpop}^2	T-dist s to est if we know σ then we use Z	$\bar D \pm t_{n-1;\alpha/2}\frac{S_D}{\sqrt n}$ n = pairs not individuals	$P(-t_{n-1;\alpha/2} < T < t_{n-1;\alpha/2}) = 1-\alpha$; $T = \frac{\bar D-\mu_D}{S_D/\sqrt n} \sim t_{n-1}$; $\bar D = \frac1n\sum_{i=1}^n D_i = \frac1n\sum_{i=1}^n(X_i-Y_i)$
μD = μ1(pop) - μ2(pop)	Big n1, n2	σ_{Dpop}^2	Zα/2 CLT s to est	$\bar D \pm z_{\alpha/2}\frac{S_D}{\sqrt n}$	$S_D^2 = \frac{1}{n-1}\sum_{i=1}^n(D_i-\bar D)^2$
σ^2_{pop}	Pop normal μ(pop)		chi^2	$\frac{\sum_{i=1}^n(X_i-\mu)^2}{\chi^2_{n;\alpha/2}} < \sigma^2 < \frac{\sum_{i=1}^n(X_i-\mu)^2}{\chi^2_{n;1-\alpha/2}}$	$P\left(\chi^2_{n-1;1-\alpha/2} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{n-1;\alpha/2}\right) = 1-\alpha$
σ^2_{pop}	Pop normal no matter size n	μ(pop)		$\frac{(n-1)S^2}{\chi^2_{n-1;\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}}$	$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^n(X_i-\bar X)^2}{\sigma^2} \sim \chi^2(n-1)$ to find S^2; S^2 = sample variance
σ_{pop}	Pop normal μ(pop)			$\sqrt{\frac{\sum_{i=1}^n(X_i-\mu)^2}{\chi^2_{n;\alpha/2}}} < \sigma < \sqrt{\frac{\sum_{i=1}^n(X_i-\mu)^2}{\chi^2_{n;1-\alpha/2}}}$	
σ_{pop}	Pop normal	μ(pop)		$\sqrt{\frac{(n-1)S^2}{\chi^2_{n-1;\alpha/2}}} < \sigma < \sqrt{\frac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}}}$	
$\frac{\sigma^2_{1\,pop}}{\sigma^2_{2\,pop}}$	Pop normal	μ1(pop) μ2(pop)		$\frac{S_1^2}{S_2^2}\frac{1}{F_{n_1-1,n_2-1;\alpha/2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}\frac{1}{F_{n_1-1,n_2-1;1-\alpha/2}}$ or $\frac{S_1^2}{S_2^2}\frac{1}{F_{n_1-1,n_2-1;\alpha/2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}F_{n_2-1,n_1-1;\alpha/2}$.	$P\left(F_{n_1-1,n_2-1;1-\alpha/2} < \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} < F_{n_1-1,n_2-1;\alpha/2}\right) = 1-\alpha$
$\frac{\sigma_{1pop}}{\sigma_{2pop}}$	Pop normal	μ1(pop) μ2(pop)		sqroot the above	

Hypothesis Testing

Decision \ Truth	H_0 is true	H_0 is false
Reject H_0	**Type I error** $P(\text{reject } H_0 \mid H_0 \text{ true}) = \alpha$	Correct Decision $P(\text{reject } H_0 \mid H_0 \text{ false}) = 1-\beta$
Not reject H_0	Correct Decision $P(\text{not reject } H_0 \mid H_0 \text{ true}) = 1-\alpha$	**Type II error** $P(\text{not reject } H_0 \mid H_0 \text{ false}) = \beta$

$$Z = \frac{\bar X-\mu_0}{\sigma/\sqrt n} \sim N(0,1).$$

$$T = \frac{\bar X-\mu_0}{S/\sqrt n} \sim t(n-1), \text{ where } S^2 \text{ is the sample variance.}$$

Then H_0 is rejected if the observed value of T, say t, satisfies

$t < -t_{n-1;\alpha/2}$ or $t > t_{n-1;\alpha/2}$. <= 2 sided
if 1 sided: reject if t < -t(n-1;alpha)

$$Z = \frac{(\bar X_1-\bar X_2)-\delta_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1).$$

(KNOWN σ_1^2, σ_2^2): NORMAL POPULATIONS OR BIG n

delta = mu1 - mu2. If unknown variance, then use S

$$T = \frac{(\bar X_1-\bar X_2)-\delta_0}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$$

$$S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

Unknown σ_1^2, σ_2^2
but known σ_1^2 == σ_2^2
Both normal, n1 n2 small

Hypothesis Test on σ^2 unknown variance

To test $H_0: \sigma^2 = \sigma_0^2$,
we can use the test statistic

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1).$$

The rejection regions for the following alternatives are given as

H_1	Rejection Region
$\sigma^2 > \sigma_0^2$	$\chi^2 > \chi^2_{n-1;\alpha}$
$\sigma^2 < \sigma_0^2$	$\chi^2 < \chi^2_{n-1;1-\alpha}$
$\sigma^2 \ne \sigma_0^2$	$\chi^2 < \chi^2_{n-1;1-\alpha/2}$ or $\chi^2 > \chi^2_{n-1;\alpha/2}$

Test Statistic: Paired Data

If $n < 30$ *is small and the differences* D_i *are normally distributed, we have the test statistic*

$$T = \frac{\bar D-\mu_{D,0}}{S_D/\sqrt n} \sim t(n-1),$$

when H_0 *is true.*
If $n \ge 30$ *is large, we have the test statistic, when* H_0 *is true,*

$$T = \frac{\bar D-\mu_{D,0}}{S_D/\sqrt n} \sim N(0,1).$$

Hypothesis Test on σ_1^2, σ_2^2 unknown means

To test $H_0: \sigma_1^2 = \sigma_2^2$,
we can use the test statistic

$$F = \frac{S_1^2}{S_2^2} \sim F(n_1-1, n_2-1).$$

The rejection regions for the following alternatives are given as

H_1	Rejection Region
$\sigma_1^2 > \sigma_2^2$	$F > F_{n_1-1,n_2-1;\alpha}$
$\sigma_1^2 < \sigma_2^2$	$F < F_{n_1-1,n_2-1;1-\alpha}$
$\sigma_1^2 \ne \sigma_2^2$	$F < F_{n_1-1,n_2-1;1-\alpha/2}$ or $F > F_{n_1-1,n_2-1;\alpha/2}$

Variance

a) var(X) >= 0
b) var(X) = E(X^2) - [E(X)]^2
c) If var(X) = 0, then P(X = mu) = 1
d) If a,b constants, var(a + bX) = b^2var(X)
e) If X and Y are independent, var(XY) = E(X^2)*E(Y^2) - [E(X)]^2 [E(Y)]^2
f) var[X + Y] = var[X] + var[Y] if X,Y independent
g) var(c) = 0 for constant c
h) var(cX) = c^2var(X), constant c
i) var(X + c) = var(X) for constant c

RULES

Commutative: E U F = F U E , E ∩ F = F ∩ E
Associative: (E U F) U G = E U (F U G), (E ∩ F)∩G = E ∩ (F∩G)
Distributive: (E U F)G = EG U FG, EF U G = (E U G)(F U G)

Combinations, Permutations

Circular arrangment: (n-1)! = n! / n
Permutation (all object distinct): nPr = n! / (n-r)!
Permutation (not all object distinct): n! / [(n1)! (n2)! ... (nk)!]
n C r = n C n-r {r = 0, 1, ...n}
n C 0 = n C n = 0
n C r = (n-1 C r) + (n-1 C r-1) {1 <= r <= n}
n C r = 0 {r < 0 or r > n}
repetition for r-combination: r+n-1 C r

Probability

(1) Equally likely outcome (2) Frequency intepretation
(3) Personal probability

Axioms of Probability:
1) For any event A, 0 <= P(A) <= 1
2) Let S be sample space, then P(S) = 1
3) For any sequence of mutually exclusive events is addition

Inclusion Exclusion

P(A U B U C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)
P(A U B) = P(A) + P(B) - P(AB) for all no matter mutually exclu or not

Conditional Probability

P(B | A) = P(A ∩ B) / P(A) P(A' | B) = 1 - P(A | B)
P(A | B') = P(A ∩ B') / P (B') = [P(A) - P(A∩B)] / [1 - P(B)]
P(A' | B') = 1 - P(A | B')
Inverse: P(A | B) = [P(B | A) * P (A)] / P(B) P(B) = P(A∩B) + P(A'∩B)

Independent

iff P(A∩B) = P(A) * P(B) or P(A | B) = P(A)
cannot check with venn diagram, only way is to prove using above
If A and B independent, so are A and B', B and A', A' and B'

Mutually exclusive

iff P(A∩B) = 0 or A∩B = {}

Pairwise Independent

iff P(Ai Aj) = P(Ai) * P(Aj) for i != j, i.e. any pair in set is independent

Mutually independent

iff for any subset you can form from original set, you satisfy pairwise independent
Mutually independent implies pairwise independence, but not the other way round

Partition

B1, ... Bn are mutually exclusive and exhaustive (i.e. B1 union ... union Bn = S)

Rule of total probability

If B1, ..., Bn is a partition of S, then for any A, P(A) = summation{ P(Bi A) }
= summation { P (Bi) * P(A | Bi) }
= P(B1 ∩ A) + P(B2 ∩ A) + ... + P(Bn ∩ A)
= P(B1) * P(A | B1) + P(B2) * P(A | B2) + ... + P(Bn) * P(A | Bn)

Bayes' Theorem

Let B1, ... Bn (causes) be a partition of S. For any event A (outcome), and k from 1 to n, P(Bk | A) = [P(Bk) * P(A | Bk)] / [P(B1)*P(A | B1) + ... + P(Bn)*P(A | Bn)]
e.g. P(B | A) = [P(Bk) * P(A | Bk)] / P(A)

Properties of PMF

$f(X_i) = P(X_i) = P(X = x_i)$

(1) f(Xi) >= 0 for every Xi (2) Summation{ f(Xi) } = 1
(3) P(X belong E) = summation of Xi belong E { f (Xi) }

Continuous Random Variable

Expectating Properties

a) E(a) = a, for constant a
b) E(X + Y) = E(X) + E(Y) regardless of independence
c) E(aX) = aE(X)
d) E(a + bX) = a + bE(X)
e) E(X*Y) = E(X) * E(Y) if X and Y are independent
For any function g(X) of random var X with PMF f(X)