Theory of Statistics

Week 0

Introducing definitions:

- **Population**: In statistical inference we seek information about some numerical characteristic of a collection of units, called population.
- **Sample**: It is often not possible to enumerate the entire population due to several constraints. In such cases, one examines a part of the population called sample. The sample must be a good representative of the population.
- **Statistical Model**: Practically it is not feasible to see the relative frequency distribution of the population phenomenon and choose F accordingly. The set of assumptions together on F together is called the statistical model.
- **Parameter**: Sometimes the form of the underlying probability distribution F is assumed to be known except for some constants, i.e., F is completely specified except the unknown constants. These unknown quantities are called parameters. To answer different questions related to this population, one needs to make inference on these parameters only. This type of inference is called *parametric inference*. An other section of inference problems are solved without assuming a particular form of F. This section is called *non-parametric inference*.
- **Statistic**: In parametric, as well as non-parametric inference, the ultimate goal is to infer a population feature, say μ based on sample realizations. A summary function of observable samples are called sample statistics.

Week 1

Let $X_1, \ldots, X_n$ be a random sample of size n from a population F. The collection of all possible values $(X_1, \ldots, X_n)$ is called the *sample space*. As sample variables are measurable functions on $\mathbb{R}$, the sample space is a subset of $\mathbb{R}^n$. Let $T(\cdot)$ be a real values function whose domain includes the sample space of $(X_1, \ldots, X_n)$ then the random variable $Y = T(X_1, \ldots, X_n)$ is called a *statistic*. The probability distribution of a statistic is called a *sampling distribution* of the statistic.

Definition 2.0.1. *Let X be a **discrete** random variable. Then the support of X, say $\mathcal{S}_X$ is the collection of points x in $\mathbb{R}$ such that $P(X = x) > 0$, i.e., $\mathcal{S}_X = \{x \in \mathbb{R} : P(X = x) > 0\}$.*

*Let X be a **continuous** random variable with CDF F_X, then the support of X, say $\mathcal{S}$ is the collection of points x in $\mathbb{R}$ such that X has a probability mass at each non-trivial neighborhood of x, i.e. $\mathcal{S}_X = \{x \in \mathbb{R} : F_X(x+h) - F_X(x+h) > 0, \forall h > 0\}$.*

Note 2.0.2. *If X is a discrete or continuous random variable with pmf or pdf f_X and support $\mathcal{S}_X$, then if $x \notin \mathcal{S}_X$, then $f_X(x) = 0$. The converse is not always true (consider the case of a pdf).*

2.1 Some important distributions

2.1.1 Discrete distributions

Bernoulli

Let $X \sim \texttt{Bernoulli}(p)$ distribution, then

$$P(X = x) = p^x(1-p)^{1-x}, \quad \mathcal{S}_X = \{0, 1\}, \quad 0 < p < 1$$

.

$$\begin{aligned} E[X] &= p.1 + (1-p).0 \\ &= p \end{aligned}$$

$$\begin{aligned} Var[X] &= E[(X - E[X])^2] \\ &= (1-p)^2 p + (0-p)^2(1-p) \\ &= p(1-p) \end{aligned}$$

$$\begin{aligned} M_X[t] &= E[e^{tX}] \\ &= e^t p + (1-p) \\ &= 1 - p + e^t p \end{aligned}$$

Binomial

Let $X \sim \texttt{Binomial}(n, p)$ distribution, then

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad \mathcal{S}_X = \{0, 1, \ldots, n\}, \quad 0 < p < 1, n \in \mathbb{N}$$

$$\begin{aligned} E[X] &= \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} i \\ &= \sum_{i=0}^{n} \binom{n-1}{i-1} np \cdot p^{i-1}(1-p)^{n-i} \\ &= np \end{aligned}$$

$$\begin{aligned} Var[X] &= E[X^2] - (E[X])^2 \\ E[X(X-1)] &= \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} x(x-1) \\ &= n(n-1)p^2 \sum_{x=0}^{n} \binom{n-2}{x-2} p^{x-2}(1-p)^{n-x} \\ E[X^2] - E[X] &= n(n-1)p^2 \\ Var[X] &= np(1-p) \end{aligned}$$

$$\begin{aligned} M_X[t] &= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= (1 - p + pe^t)^n \end{aligned}$$

Poisson

Let $X \sim \texttt{Poisson}(\lambda)$ distribution, then

$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad \mathcal{S}_X = \{0, 1, \dots\}, \quad \lambda > 0$$

$$\begin{aligned} E[X] &= \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!}\lambda \\ &= \lambda e^{\lambda} e^{-\lambda} \\ &= \lambda \end{aligned}$$

$$\begin{aligned} Var[X] &= E[X^2] - (E[X])^2 \\ E[X(X-1)] &= \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^{x-2}}{(x-2)!}\lambda^2 \\ E[X^2] - E[X] &= \lambda^2 \\ Var[X] &= \lambda \end{aligned}$$

$$\begin{aligned} M_X[t] &= \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x e^{tx}}{x!} \\ &= \sum_{x=0}^{\infty} \frac{e^{-\lambda}(\lambda e^t)^x}{x!} \frac{e^{-\lambda e^t}}{e^{-\lambda e^t}} \\ &= \frac{e^{-\lambda}}{e^{-\lambda e^t}} \\ &= e^{\lambda(e^t - 1)} \end{aligned}$$

Geometric

Let $X \sim \texttt{Geometric}(p)$ distribution, then
$P(X = x) = (1-p)^{x-1}p, \quad \mathcal{S}_X = \{1, \dots\}, \quad 0 < p < 1$

$$\begin{aligned} E[X] &= \sum_{x=1}^{\infty} (1-p)^{x-1}px \quad \text{(an AGP)} \\ &= \frac{1}{p} \end{aligned}$$

$$\begin{aligned} Var[X] &= E[X^2] - (E[X])^2 \\ E[X(X-1)] &= \sum_{x=1}^{\infty} (1-p)^{1-x}px(x-1) \quad \text{(another AGP)} \\ &= \frac{2(1-p)}{p^2} \\ Var[X] &= \frac{1-p}{p^2} \\ M_X[t] &= \sum_{x=1}^{\infty} (1-p)^{x-1}p \cdot e^{tx} \quad \text{(a GP)} \\ &= \frac{pe^t}{1-(1-p)e^t} \end{aligned}$$

2.1.2 Continuous distributions

Uniform

Let $X \sim \texttt{Uniform}(\alpha, \beta)$ distribution, then
$f_X(x) = \frac{1}{\beta-\alpha}, \alpha < x < \beta, \mathcal{S}_X = [\alpha, \beta], \alpha, \beta \in \mathbb{R}, \beta > \alpha$

$$\begin{aligned} E[X] &= \int_{\alpha}^{\beta} \frac{x}{(\beta-\alpha)}dx \\ &= \frac{\beta+\alpha}{2} \end{aligned}$$

$$\begin{aligned} E[X(X-1)] &= \int_{\alpha}^{\beta} \frac{x(x-1)}{(\beta-\alpha)}dx \\ &= \frac{\beta^2+\alpha\beta+\alpha^2}{3} \\ Var[X] &= E[X^2] - (E[X])^2 \\ &= \frac{(\beta-\alpha)^2}{12} \end{aligned}$$

$$\begin{aligned} M_X[t] &= \int_{\alpha}^{\beta} \frac{e^{tx}}{(\beta-\alpha)}dx \\ &= \frac{e^{t\beta} - e^{t\alpha}}{t(\beta-\alpha)} \end{aligned}$$

Gamma

Let $X \sim \texttt{Gamma}(\alpha, \beta)$, then
$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}, \quad x > 0, \mathcal{S}_X = [0, \infty), \alpha > 0, \beta > 0$

Note 2.1.1. *The Gamma Integral:* $\int_0^\infty x^{\alpha-1}e^{-x}dx = \Gamma(\alpha)$

$$\begin{aligned} E[X] &= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)}x^\alpha e^{-\beta x}dx \\ &= \int_0^\infty \frac{(\beta x)^\alpha e^{-\beta x}}{\Gamma(\alpha)}dx \\ &= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \\ &= \frac{\alpha}{\beta} \\ E[X^2] &= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha+1}e^{-\beta x}dx \\ &= \frac{1}{\beta\Gamma(\alpha)}\int_0^\infty (\beta x)^{\alpha+1}e^{-\beta x}dx \\ &= \frac{\alpha^2+\alpha}{\beta^2} \\ Var[X] &= \frac{\alpha}{\beta^2} \end{aligned}$$

$$
\begin{aligned}
M_X[t] &= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \cdot e^{tx} dx \\
&= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\beta-t)x} dx \\
&= \frac{1}{(1-t/\beta)^\alpha} \quad \text{(change of variable } (\beta-t)x = y\text{)}
\end{aligned}
$$

Exponential

Let $X \sim$ `Exponential`(λ): Special case of the Gamma distribution with $\alpha = 1, \beta = \lambda$

Beta

Let $X \sim$ `Beta`(α, β), then
$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, 0 < x < 1, \mathcal{S}_X = [0,1], \alpha > 0, \beta > 0$

Note 2.1.2. *The Beta Integral:* $\int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx = Beta(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

$$
\begin{aligned}
E[X] &= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha}(1-x)^{\beta-1} dx \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} Beta(\alpha+1,\beta) \\
&= \frac{\alpha}{\alpha+\beta} \\
E[X^2] &= \frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)} \\
Var[X] &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}
\end{aligned}
$$

Normal

Let $X \sim$ `Normal`(μ, σ^2) distribution, then
$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, x \in \mathcal{S}_X = \mathbb{R}, \mu \in \mathbb{R}, \sigma > 0$

$$
\begin{aligned}
E[X] &= \mu \\
Var[X] &= \sigma^2 \\
M_X[t] &= e^{\mu t + \sigma^2 t^2/2}
\end{aligned}
$$

Cauchy

Let $X \sim$ `Cauchy`(μ, σ) distribution, then
$f_X(x) = \frac{1}{\pi\sigma}\left[1 + \left(\frac{x-\mu}{\sigma}\right)^2\right]^{-1}, x \in \mathcal{S}_X = \mathbb{R}, \mu \in \mathbb{R}, \sigma > 0$

The expectation does not exist for this distribution. For instance consider $\mu = 0, \sigma = 1$
$E[X^+] = \int_0^\infty \frac{x}{\pi(1+x^2)} dx = \infty$
$E[X^-] = -\int_{-\infty}^0 \frac{x}{\pi(1+x^2)} dx = \infty$
Hence $E[X]$ is not defined.

Chi-squared

Let $X_i \overset{\text{iid}}{\sim}$ `Normal`$(0,1)$ distribution, $i = 1, \ldots, n$, then $T = \sum_{i=1}^n X_i^2$ follows a `Chi-squared` distribution with degrees of freedom n, i.e. $T \sim \chi_n^2$ and,
$f_T(t) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}, \mathcal{S}_X = [0,\infty), n \in \mathbb{N}$
$M_X[t] = (1-2t)^{-n/2}$. Now to show that if $X_i \sim \mathcal{N}(0,1)$ then $T \sim \chi_n^2$

$$
\begin{aligned}
Y &= X_i^2 \\
E[e^{ty}] &= E[e^{tx^2} \mid X \in \mathcal{N}(0,1)] \\
&= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot e^{tx^2 dx} \\
&= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2(1-2t)/2} dx \\
&= \frac{1}{\sqrt{1-2t}}
\end{aligned}
$$

Since the moment-generating-function (MGF) is unique for each distribution, hence the required result is arrived at.

2.2 Some properties (1)

2.2.1 Additive Properties

1. Let $X_i \overset{\text{ind}}{\sim}$ `Binomial`(n_i, p) for $i = 1, \ldots, k$, then $T = \sum_{i=1}^k X_i$ follows `Binomial`$(\sum_i n_i, p)$.

   Proof. $M_T[t] = E[e^{tT}] = E[e^{t\sum_i X_i}] = E[\prod_i e^{tX_i}] = \prod_i E[e^{tX_i}]$.

   The last statement holds because of independence, hence $M_T[t] = \prod_i E[e^{tX_i}] = \prod_i M_{X_i}[t] = \prod_i (1-p+pe^t)^{n_i} = (1-p+pe^t)^{\sum_i n_i}$, which characterizes `Binomial`$(\sum_i n_i, p)$. ☐

2. Let $X_i \overset{\text{ind}}{\sim}$ `Poisson`(λ_i) for $i = 1, \ldots, k$, then $T = \sum_{i=1}^k X_i$ follows `Poisson`$(\sum_i \lambda_i)$.

   Proof. $M_T[t] = \prod_i E[e^{tX_i}] = \prod_i e^{\lambda_i(e^t-1)} = e^{\sum_i \lambda_i(e^t-1)}$ which characterizes `Poisson`$(\sum_i \lambda_i)$ ☐

3. Let $X_i \overset{\text{ind}}{\sim} \mathcal{N}(\mu_i, \sigma_i^2)$ for $i = 1, \ldots, k$, then $T = \sum_{i=1}^k X_i$ follows $\mathcal{N}(\sum_i \mu_i, \sum_i \sigma_i^2)$.

   Proof. $M_T[t] = \prod_i E[e^{tX_i}] = \prod_i e^{\mu_i t + \sigma_i^2 t^2/2} = e^{\sum_i \mu_i \cdot t \sum_i \sigma_i^2 \cdot t^2}$ which characterizes $\mathcal{N}(\sum_i \mu_i, \sum_i \sigma_i^2)$ ☐

4. Let $X_i \overset{\text{ind}}{\sim}$ `Gamma`(α_i, β) for $i = 1, \ldots, k$, then $T = \sum_{i=1}^k X_i$ follows `Gamma`$(\sum_i \alpha_i, \beta)$.

Proof. $M_T[t] = \prod_i E[e^{tX_i}] = \prod_i \frac{1}{(1-t/\beta)_i^{\alpha}} = \frac{1}{(1-t/\beta)^{\sum_i \alpha_i}}$ which characterizes $\texttt{Gamma}(\sum_i \alpha_i, \beta)$. □

5. Let $X_i \overset{\text{ind}}{\sim} \chi^2_{n_i}$ for $i = 1, \dots, k$, then $T = \sum_{i=1}^k X_i$ follows χ^2_N where $N = \sum_i n_i$.

   Proof. Since $X_i \overset{\text{ind}}{\sim} \chi^2_n$ implies there exist $Y_{i1}, Y_{i2}, \dots, Y_{in_i}$ such that $Y_{ij} \overset{\text{ind}}{\sim} \mathcal{N}(0,1)$ for $j = 1, \dots, n_i$. $T = \sum_i X_i = \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij}^2$ wherein all the Y_{ij}'s follow $\mathcal{N}(0,1)$ and are independent of one other. Hence $T = \sum_{\ell=1}^N Y_\ell^2$, and hence T follows χ^2_N. □

Some other properties are,

⇒ Let $X \sim \mathcal{N}(\mu, \sigma^2)$ distribution, then $T = aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

$$\begin{aligned} M_T[t] &= E[e^{tT}] \\ &= E[e^{t(aX+b)}] \\ &= E[e^{(at)X+tb}] \\ &= e^{\mu(at)+bt+a^2t^2\sigma^2/2} \end{aligned}$$

This characterizes a Normal distribution with mean $a\mu + b$ and variance $a^2\sigma^2$.

⇒ Let $X \sim \texttt{Gamma}(\alpha, \beta)$ distribution, then $T = aX \sim \texttt{Gamma}(\alpha, \beta/a)$.

$$\begin{aligned} M_T[t] &= E[e^{(at)x}] \\ &= \frac{1}{(1 - t/(\beta/a))^{\alpha}} \end{aligned}$$

This characterizes the distribution $\texttt{Gamma}(\alpha, \beta/a)$.

⇒ Let $X \sim \texttt{Beta}(n/2, m/2)$ distribution, then $T = mX/\{n(1-X)\} \sim F_{n,m}$.

Given the function for $T \to X = \frac{nT}{nT+m}$ $[X = g^{-1}(T)]$.
$f_T(t) = f_X(g^{-1}(t)) \cdot \frac{\partial X}{\partial T}$
$= \frac{\Gamma((n+m)/2)}{\Gamma(n/2)\Gamma(m/2)} \left(\frac{nT}{nT+m}\right)^{n/2-1} \left(\frac{m}{nT+m}\right)^{m/2-1} \cdot \frac{mn}{(nT+m)^2}$
which upon minor simplification gives the $F_{n,m}$ distribution.

⇒ Let $X \sim \texttt{Uniform}(0,1)$, and $\alpha > 0$ then $T = X^{1/\alpha} \sim \texttt{Beta}(\alpha, 1)$.

$X = T^{\alpha}$ hence $f_T(t) = f_X(t^{\alpha}) \cdot \frac{\partial X}{\partial T}$
$f_T(t) = \alpha x^{\alpha-1}$ under the domain of $T \in (0,1)$ which is nothing but the desired $\texttt{Beta}(\alpha, 1)$.

⇒ Let $X \sim \texttt{Cauchy}(0,1)$ distribution, then $T = 1/(1+X^2) \sim \texttt{Beta}(0.5, 0.5)$.

The range of T is $(0,1)$ and $X = \sqrt{\frac{1}{T} - 1}$
$f_T(t) = \frac{T}{\pi} \cdot \left|\frac{\partial X}{\partial T}\right| = \frac{T}{\pi} \cdot \frac{1}{T^2\left(\frac{1}{T}-1\right)^{1/2}} = \frac{1}{\pi} T^{0.5}(1-T)^{0.5}$
which is the $\texttt{Beta}(0.5, 0.5)$ distribution.

⇒ Let $X \sim \texttt{Uniform}(0,1)$ distribution then $T = -2\log X \sim \chi^2_2$.

The range of T is $(0,\infty), X = e^{-t/2}$, hence $f_T(t) = \frac{e^{-t/2}}{2}$.

⇒ Let X be distributed as some absolutely continuous distribution with cdf G_X then $T = G_X(X) \sim \texttt{Uniform}(0,1)$.

CDF for T, $F_T(t) = P(T \le t) = P(G_X(X) \le t) = P(X \le G_X^{-1}(t)) = G_X(G_X^{-1})(t) = t$. To characterize it as the uniform distribution, the range of T and the fact that it is absolutely continuous has to be used to establish the exactness of the distribution.

2.3 Multivariate distributions

Suppose X_I is an absolutely continuous (or, discrete) random variable with pdf (or, pmf) f_i, $i = 1, \dots, n$ and $X_1, \dots, X_n$ are mutually independent then the multivariate distribution of the random vector $\mathbf{X} = (X_1, \dots, X_n)$ has the joint pdf (or pmf) $f_{\mathbf{X}}$ where

$$f_{\mathbf{X}}(x_1, \dots, x_n) = f_1(x_1) \times \cdots \times f_n(x_n) \forall \mathbf{x} \in \mathbb{R}^n$$

Notation: X and Y are independent $\to X \amalg Y$
Independence is the only case in which total distribution is captured just by marginal distributions.

2.3.1 Characterizing the joint distribution

- **CDF**: $F_{X,Y}(x,y) = P(X \le x, Y \le y)$
- **Characteristic function**: $C_{\mathbf{X}}(\mathbf{x}) = E[e^{it'\mathbf{X}}]$
- **Discrete PMF**: $f_{X,Y}(x,y) = P(X = x, Y = y)$
- **Absolutely continuous PDF** $f_{X,Y}(x,y) = \lim_{h_1\to\infty, h_2\to\infty} \frac{P(x-h_1 \le X \le x+h_1, y-h_2 \le Y \le y+h_2)}{4h_1h_2}$

2.3.2 Marginal distribution

Let $\mathbf{X}$ be a k-dimensional random vector with CDF $F_{\mathbf{X}}$, then the marginal CDF of the j-th component of $\mathbf{X}$, X_j is

$$F_{X_j}(x) = F_{\mathbf{X}}(\infty, \ldots, \underbrace{x}_{\text{j-th}}, \infty, \ldots, \infty), x \in \mathbb{R}$$

Simply put, consider the case of a bi-variate case,
$F_X(x) = F_{X,Y}(x, y) = \lim_{y\to\infty} F_{X,Y}(x, y)$
Also, for PMF and PDF, consider this,

PMF: $f_X(x) = \sum_{y\in\mathcal{S}} f_{X,Y}(x, y)$
PDF: $f_X(x) = \int_y f_{X,Y}(x, y)dy$

2.3.3 Conditional distribution

Let $(\mathbf{X}, \mathbf{Y})'$ be a discrete random variable, then the conditional distribution of X given $\mathbf{Y} = \mathbf{y}$ is,
$f_{\mathbf{X}\,|\,\mathbf{Y}=\mathbf{y}}(\mathbf{x}) = P(\mathbf{X} = \mathbf{x}\,|\,\mathbf{Y} = \mathbf{y}) = \frac{P(\mathbf{X}=\mathbf{x},\mathbf{Y}=\mathbf{y})}{P(\mathbf{Y}=\mathbf{y})} = \frac{f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})}$
For $(X, Y)'$ be absolutely continuous random variable,
$F_{X\,|\,Y=y}(x) = \lim_{h\to\infty} P(X \le x\,|\,y - h < Y \le y + h)$
and
$f_{X\,|\,Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$

2.3.4 Some Derived Distributions

F-distribution

Let $X \sim \chi^2_{n_1}, Y \sim \chi^2_{n_2}$ and X and Y are independently distributed, then $F = \frac{n_2 X}{n_1 Y}$ follows an F distribution with d.f. n_1 and n_2 notationally $F \sim F_{n_1,n_2}$ and the pdf of F, f_F is given by

$$f_F(x) = \frac{\Gamma((n_1 + n_2)/2)}{\Gamma(n_1/2)\Gamma(n_2/2)} \left(\frac{n_1}{n_2}\right)^{n_1/2} x^{n_1/2-1} \left(1 + \frac{n_1}{n_2}x\right)^{-(n_1+n_2)/2}$$

where $x > 0, \mathcal{S}_X = [0, \infty)$.

Let $F = n_2X/n_1Y$ and let $T = 1/n_1Y$ which gives $X = F/n_2T$ and $Y = 1/n_1T$.
$f_{F,T}(f, t) = f_{X,Y}(f/n_2t, 1/n_1t).|J|$ where $J = \begin{vmatrix} \partial x/\partial f & \partial x/\partial t \\ \partial y/\partial f & \partial y/\partial t \end{vmatrix} = -\frac{1}{n_1n_2t^3}$ and since X, Y are independent the distribution $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$.
$f_{F,T}(f, t) = \frac{f^{n_1/2-1}e^{-\frac{1}{2t}(f/n_2+1/n_1)}}{\Gamma(n_1/2)\Gamma(n_2/2)2^{(n_1+n_2)/2}t^{\{(n_1+n_2)/2+1\}}n_1^{n_2/2}n_2^{n_1/2}}$
$f_F(f) = \int_0^\infty f_{F,T}(f, t).dt$
The relevant integral becomes
$\int_0^\infty \left(\frac{1}{t}\right)^{(n_1+n_2)/2+1} e^{-\frac{1}{2t}(f/n_2+1/n_1)}dt$
A change of variable $t = 1/u$ gives the following integral to be $\frac{\Gamma((n_1+n_2)/2)}{\beta^{(n_1+n_2)/2}}$ where $\beta = -\frac{1}{2}\left(\frac{f}{n_2} + \frac{1}{n_1}\right)$. A simple substitution gives the desired distribution in functional form.

t-distribution

Let $X \sim \mathcal{N}(0, 1), Y \sim \chi^2_n$ and X and Y are independently distributed. Then $W = X/\sqrt{Y/n}$ follows t-distribution with d.f. n, notationally $W \sim t_n$ and the pdf of W, f_W is given by

$$f_W(x) = \frac{\Gamma((n+1)/2)}{\sqrt{\pi n}\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}, \quad x \in \mathcal{S}_X = \mathbb{R}$$

Let $W = X/\sqrt{Y/n}, T = \sqrt{Y}$ which gives $Y = T^2$ and $X = TW/\sqrt{n}$
$f_{W,T}(w, t) = f_{X,Y}(tw/\sqrt{n}, t^2) \cdot |J|$ where $J = \begin{vmatrix} \partial x/\partial w & \partial x/\partial t \\ \partial y/\partial w & \partial y/\partial t \end{vmatrix} = 2t^2/\sqrt{n}$ and since X, Y are independent the distribution $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$.
$f_{W,T}(w, t) = \frac{2e^{-\frac{t^2}{2}(w^2/n+1)}t^n}{\sqrt{2n\pi}2^{n/2}\Gamma(n/2)}$
$f_W(w) = \int_0^\infty f_{W,T}(w, t).dt$ Hence the above when substituted in the integral and changing the variable $t^2 = u$ gives the desired functional form of the t-distribution.

2.3.5 Multivariate Normal Distribution

$\mathbf{X} \sim \mathcal{N}_k(\boldsymbol{\mu}, \Sigma)$ where $\boldsymbol{\mu} \in \mathbb{R}^k$ and Σ is a positive definite matrix, if and only if, $\mathbf{a}^T\mathbf{X} \sim \mathcal{N}(\mathbf{a}^T\boldsymbol{\mu}, \mathbf{a}^T\Sigma\mathbf{a}), \forall \mathbf{a} \in \mathbb{R}^k$.
Is it a valid distribution?
Consider $M_{\boldsymbol{X}}[\boldsymbol{t}] = E[e^{\boldsymbol{t}^T\boldsymbol{X}}] = M_{\boldsymbol{Y}}[1]$ where $\boldsymbol{t}^T\boldsymbol{X} = \boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{t}^T\boldsymbol{\mu}, \boldsymbol{t}^T\Sigma\boldsymbol{t})$
Hence $M_{\boldsymbol{X}}[\boldsymbol{t}] = e^{\boldsymbol{t}^T\boldsymbol{\mu}+\boldsymbol{t}^T\Sigma\boldsymbol{t}/2}$. Which implies that the distribution is valid.
Now, let $\boldsymbol{X} \sim \mathcal{N}_k(\boldsymbol{\mu}, \Sigma)$

$$E[\mathbf{X}] = \begin{bmatrix} E[X_1] \\ \vdots \\ E[X_k] \end{bmatrix} = \boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)$$

To get the marginal distribution of X_i take $\boldsymbol{a}$'s to be $\boldsymbol{e}_i$'s.

For a multivariate distribution, the variance is defined as $Var[\boldsymbol{X}] = E[(\boldsymbol{X}-E[\boldsymbol{X}])(\boldsymbol{X}-E[\boldsymbol{X}])^T]$, which yields the *Variance-Covariance matrix.*

$$= E\left[\begin{pmatrix} X_1 - \mu_1 \\ \vdots \\ X_k - \mu_k \end{pmatrix}((X_1 - \mu_1)\ldots(X_k - \mu_k))\right]$$
$$= E\begin{bmatrix} (X_1 - \mu_1)^2 & \ldots & (X_1 - \mu_1)(X_k - \mu_k) \\ \vdots & & \vdots \\ (X_1 - \mu_1)(X_k - \mu_k) & \ldots & (X_k - \mu_k)^2 \end{bmatrix}$$
$$= \begin{bmatrix} Var[X_1] & Cov[X_1, X_2] & \ldots & Cov[X_1, X_k] \\ \vdots & \vdots & & \vdots \\ Cov[X_1, X_k] & Cov[X_2, X_k] & \ldots & Var[X_k] \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1k} \\ \vdots & \vdots & & \vdots \\ \sigma_{k1} & \sigma_{k2} & \dots & \sigma_{kk} \end{bmatrix}$$

Hence $E[\boldsymbol{X}] = \boldsymbol{\mu}$, $Var[\boldsymbol{X}] = \Sigma$.

If the focus is on cases when Σ is positive definite then the pdf

$f_{\boldsymbol{X}}(\boldsymbol{x}) = (2\pi)^{-n/2}|\Sigma|^{-1/2}\exp\{\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}$, $\boldsymbol{\mu} \in \mathbb{R}^k$.

For the bi-variate case, five parameters define the distribution and is given by

$$f_{\boldsymbol{X}}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2+\left(\frac{y-\mu)y}{\sigma_y}\right)^2-2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\}}$$

where $x, y \in \mathbb{R}, \mu_x, \mu_y \in \mathbb{R}, \sigma_x, \sigma_y > 0, 0 \le \rho \le 1$.

2.4 Some properties (2)

$\Rightarrow$ [*Box-Muller transformation*] Let $U_i \overset{\text{iid}}{\sim}$ `Uniform`$(0,1), i = 1,2$. Consider the transformations $Z_1 = \sqrt{-2\log U_1}\cos(2\pi U_2)$ and $Z_2 = \sqrt{-2\log U_1}\sin(2\pi U_2)$. Then $(Z_1, Z_2)^T \sim \mathcal{N}_2(\boldsymbol{0}, I)$ where I is the identity matrix.

The range of Z_1, Z_2 is $(0,\infty)$ and $U_1 = e^{-\frac{Z_1^2+Z_2^2}{2}}$ and $U_2 = \frac{1}{2\pi}\tan^{-1}\left(\frac{Z_1}{Z_2}\right)$ and hence $f_{Z_1,Z_2}(z_1,z_2) = \begin{vmatrix} e^{-\frac{z_1^2+z_2^2}{2}}(-z_1) & e^{-\frac{z_1^2+z_2^2}{2}}(-z_2) \\ \frac{z_2}{2\pi(z_1^2+z_2^2)} & \frac{-z_1}{2\pi(z_1^2+z_2^2)} \end{vmatrix} = \frac{e^{-z_1^2/2}}{\sqrt{2\pi}} \cdot \frac{e^{-z_2^2/2}}{\sqrt{2\pi}}$

which means that individually Z_1 and Z_2 are distributed according to the standard normal distribution and hence cumulatively the desired statement holds.

$\Rightarrow$ Let $X_i \overset{\text{iid}}{\sim}$ `Gamma`$(\alpha_i, \beta), i = 1,2$. Consider the transformation $Z = X_1/(X_1 + X_2)$. Then $Z \sim$ `Beta`(α_1, α_2).

Let $T = X_1 + X_2$, which means that $X_1 = TZ, X_2 = T(1-Z)$ and the range of T is $(0,\infty)$ while the range of Z is $(0,1)$.

$f_{Z,T}(z,t) = f_{X_1,X_2}(tz, t(1-z)).|J|$ where $J = \begin{vmatrix} \partial x_1/\partial z & \partial x_1/\partial t \\ \partial x_2/\partial z & \partial x_2/\partial t \end{vmatrix} = t$. Hence $f_{Z,T}(z,t) = \frac{\beta^{\alpha_1+\alpha_2}z^{\alpha_1-1}(1-z)^{\alpha_2-1}t^{\alpha_1+\alpha_2-1}e^{-\beta t}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}$.

Since $f_Z(z) = \int_0^\infty f_{Z,T}(z,t).dt$ the desired `Beta` distribution is arrived at.

$\Rightarrow$ Let (X, Y) is jointly distributed as $\mathcal{N}_2(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$. Then the marginal distribution of X is $\mathcal{N}(\mu_x, \sigma_x^2)$. Also the conditional distribution of Y given $X = x$ is $\mathcal{N}(\mu_y + \rho\sigma_y(x-\mu_x)/\sigma_x, \sigma_y^2(1-\rho^2))$.

Since $(X, Y)^T \sim \mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$ where $\boldsymbol{\mu}^T = (\mu_1, \mu_2)$ and $\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$. This implies that $\boldsymbol{a}^T\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}(\boldsymbol{a}^T\boldsymbol{\mu}, \boldsymbol{a}^T\Sigma\boldsymbol{a})$. Take $\boldsymbol{a} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ to get that the marginal distribution of X is $\mathcal{N}(\mu_x, \sigma_x^2)$.

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2+\left(\frac{y-\mu)y}{\sigma_y}\right)^2-2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\}}}{\frac{1}{\sqrt{2\pi}\sigma_x}e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}}$$

which reduces to exactly as desired.

$\Rightarrow$ Let $X_i \overset{\text{iid}}{\sim}$ `Bernoulli`$(p), i = 1, \dots, n$. Then the conditional distribution of $\boldsymbol{X}$ given $\overline{X}_n = y$ is free of p.

$P(\boldsymbol{X} = \boldsymbol{x} \mid \overline{X}_n = y) = \frac{P(\boldsymbol{X}=\boldsymbol{x}, \overline{X}_n = y)}{P(\overline{X}_n = y)}$

$P(\overline{X}_n = y) = \binom{n}{ny} p^{ny}(1-p)^{n(1-y)}$, while

$P(\boldsymbol{X} = \boldsymbol{x}, \overline{X}_n = y) = p^{ny}(1-p)^{n(1-y)}$ OR 0

Hence $P(\boldsymbol{X} = \boldsymbol{x} \mid \overline{X}_n = y) = \frac{1}{\binom{n}{ny}}$ OR 0

$\Rightarrow$ Let $X_i \overset{\text{iid}}{\sim}$ `Poisson`$(\lambda), i = 1, \dots, n$. Then the conditional distribution of $\boldsymbol{X}$ given $\overline{X}_n = y$ is free of λ.

$P(\boldsymbol{X} = \boldsymbol{x} \mid \overline{X}_n = y) = \frac{P(\boldsymbol{X}=\boldsymbol{x}, \overline{X}_n = y)}{P(\overline{X}_n = y)}$

$P(\boldsymbol{X} = \boldsymbol{x} \mid \overline{X}_n = y) = \frac{\frac{e^{-n\lambda}\lambda^{ny}}{(x_1!x_2!\dots x_n!)}}{\sum_{i_1,i_2,\dots,i_n \mid x_{i_1}+\dots+x_{i_n}} \frac{e^{-n\lambda}\lambda^{ny}}{x_{i_1}!\dots x_{i_n}!}}$ OR 0

and hence it is independent of λ.

Week 2

3.1 Some important statistics and their sampling distribution

3.1.1 Sample Mean

Let $X_1, \ldots, X_n$ be a random sample from some distribution F. Then $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is called the sample mean.

Properties of sample mean

(a) Let $X_1, \ldots, X_n$ be a random sample from some distribution F with expectation μ and finite variance σ^2, then $E[\overline{X}_n] = \mu$ and $Var[\overline{X}_n] = \sigma^2/n$.

Proof. $E[\overline{X}_n] = \frac{1}{n}E[\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \mu$.

Now for the variance, wherein $\mu_2 = E\left[X_i^2\right] \leq \infty$,

$$\begin{aligned} Var\left[\overline{X}_n\right] &= E\left[\left(\overline{X}_n - E[\overline{X}_n]\right)^2\right] \\ &= E\left[\overline{X}_n^2\right] - \left(E\left[\overline{X}_n\right]\right)^2 \end{aligned}$$

$$\begin{aligned} E\left[\overline{X}_n^2\right] &= E\left[\frac{1}{n^2}\left(\sum_{i=1}^{n} X_i\right)^2\right] \\ &= \frac{1}{n^2}E\left[\sum_{i} E[X_i^2] + \sum_{i\neq j} E[X_iX_j]\right] \\ &= \frac{1}{n^2}E\left[n\mu_2 + \sum_{i\neq j} E[X_i]E[X_j]\right] \\ &= \frac{\mu_2}{n} + \frac{(n-1)\mu^2}{n} \end{aligned}$$

$$Var\left[\overline{X}_n\right] = \left(\frac{\mu_2-\mu^2}{n}\right)^2 = Var[X_i]/n = \sigma^2/n \qquad \square$$

(b) Let $X_1, \ldots, X_n$ be a random sample from some distribution F with expectation μ and finite variance σ^2, then $\overline{X}_n$ is the *best linear unbiased estimator* BLUE of μ.

What is the BLUE?

A linear estimator is nothing but, $\sum_i \ell_i X_i = \ell(\mathbf{X})$. A linear unbiased estimator would be $\sum_i \ell_i X_i$ s.t. $E[\ell(\mathbf{X})] = \mu$, this implies $\sum_i \ell_i = 1$.

Hence the best linear unbiased estimator is the one in which $Var[\ell(\mathbf{X})]$ must be minimum among all other linear unbiased estimators.

$Var[\ell(\mathbf{X})] = Var[\ell_1 X_1 + \cdots + \ell_n X_n] = \sum_{i=1}^{n} \ell_j^2\sigma^2$ is minimum under the constraint that $\sum_i \ell_i = 1/n$. Solving using Lagrangian multiplier, the desired result is obtained.

(c) Let $X_1, \ldots X_n$ be a random sample from $\mathcal{N}(\mu, \sigma^2)$ distribution. Then the sampling distribution of $\overline{X}_n$ is $\mathcal{N}(\mu, \sigma^2/n)$.

Use the additive properties of normal distribution, i.e., $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2) \Rightarrow \sum_i X_i \sim \mathcal{N}(n\mu, n\sigma^2)$, and use the fact that if $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2) \Rightarrow aY + b \sim \mathcal{N}(a\mu_Y + b, a^2\sigma_Y^2)$.

3.1.2 Sample variance

Let $X_1, \ldots X_n$ be a random sample from some distribution F. Then

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2$$

is called the sample variance.

Properties of sample variance

(a) Let $X_1, \ldots X_n$ be a random sample from some distribution F with expectation μ and variance σ^2. Then $E\left[S_n^2\right] = \frac{n-1}{n}\sigma^2$

Proof.

$$E\left[S_n^2\right] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} E[X_i^2] - E\left[\overline{X}_n^2\right]$$
$$= \left(\sigma^2 + \mu^2\right) - (\mu^2 + \sigma^2/n)$$
$$= \sigma^2\left(\frac{n-1}{n}\right)$$

□

This means $E\left[S_n^{*2}\right] = \sigma^2$ where $S_n^{*2} = \frac{n}{n-1}S_n^2$ is called the unbiased sample variance.

(b) Let $X_1, \dots X_n$ be a random sample from $\mathcal{N}(\mu, \sigma^2)$ distribution. Then the sampling distribution of nS_n^2 is $\sigma^2\chi^2_{n_1}$. Further the sampling distribution of $\overline{X}_N$ is $\mathcal{N}(\mu, \sigma^2/n)$, and these two are independent of each other.

Proof. Consider the orthogonal transformation $(X_1, X_2, \dots, X_n) \longrightarrow (W_1, W_2, \dots, W_n)$, wherein

$$W_1 = \frac{1}{\sqrt{n}}\left[\frac{X_1-\mu}{\sigma} + \cdots + \frac{X_n-\mu}{\sigma}\right]$$
$$W_2 = a_{21}\left(\frac{X_1-\mu}{\sigma}\right) + \cdots + a_{2n}\left(\frac{X_n-\mu}{\sigma}\right)$$
$$\vdots$$
$$W_n = a_{n1}\left(\frac{X_1-\mu}{\sigma}\right) + \cdots + a_{nn}\left(\frac{X_n-\mu}{\sigma}\right)$$

such that

$$\begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_n \end{bmatrix} = \underbrace{\begin{bmatrix} 1/\sqrt{n} & \dots & 1/\sqrt{n} \\ a_{21} & \dots & a_{2n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}}_{A} \begin{bmatrix} (X_1-\mu)/\sigma \\ (X_2-\mu)/\sigma \\ \vdots \\ (X_n-\mu)/\sigma \end{bmatrix}$$

s.t. $A^TA = I$. Such an A exists as we can have $(n-1)$ orthogonal unit vectors in that vector space in which $\frac{1}{\sqrt{n}}\mathbf{1}$ resides.

$\boldsymbol{W} = A(\boldsymbol{X} - \mu\mathbf{1})/\sigma$. This means $f_{\boldsymbol{W}}(\boldsymbol{w}) = f_{\boldsymbol{W}}(\sigma A^T\boldsymbol{w} + \mu\mathbf{1}) \cdot J$.

Consider

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{X_1}(x_1)\cdot\dots\cdot f_{X_n}(x_n)$$
$$= \frac{1}{(2\pi)^{n/2}\sigma^n} e^{\frac{-1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2}$$

$\sum_{i=1}^n (x_i - \mu)^2 = ||\boldsymbol{x} - \mu\mathbf{1}||^2$ which is $\sigma^2(A^T\boldsymbol{w})^T(A^T\boldsymbol{w}) = \boldsymbol{w}^T\boldsymbol{w} = ||\boldsymbol{w}||^2 = \sum_{i=1}^n w_i^2$

And the jacobian J is simple $|\sigma A^T| = \sigma^n$.

This gives $f_{\boldsymbol{W}}(\boldsymbol{w}) = \frac{1}{(2\pi)^{n/2}} e^{\frac{-1}{2}\sum_i w_i^2}$, which in turn gives $w_i \sim \mathcal{N}(0,1)$ for all $i = 1, \dots, n$.

By the definition of *Chi-square* distribution, $\sum_{i=2}^n w_i^2 \sim \chi^2_{n-1}$

$$\sum_{i=2}^n w_i^2 = \sum_{i=1}^n w_i^2 - w_1^2$$
$$= \frac{1}{\sigma^2}\left[\sum_{i=1}^n (x_i-\mu)^2 - \frac{1}{n}(n\overline{x}_n - n\mu)\right]$$
$$= \frac{1}{\sigma^2}\left[\sum_{i=1}^n x_i^2 - 2\mu\sum_{i=1}^n x_i + n\mu^2 - n\overline{x}_n^2 - n\mu^2 + 2n\overline{x}_n\mu\right]$$
$$= \frac{nS_n^2}{\sigma^2}$$

Now, completing the argument about the independence of $\overline{X}_n$ and S_n^2. W_i's are independent and $\overline{X}_n$ can be presented as a pure function of W_1: $\overline{X}_n = \frac{\sigma}{\sqrt{n}}W_1 + \mu$ and S_n^2 is shown as a function of $W_2, \dots, W_n$, hence $\overline{X}_n$ and S_n^2 are independent. □

Corollary 3.1.0.1. $\frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} \sim t_{n-1}$.

3.1.3 Sample correlation

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a bi-variate random sample from some distribution F, then the sample correlation coefficient is

$$r_{x,y} = \frac{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)}{\sqrt{\left\{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2\right\}\left\{\frac{1}{n}\sum_{i=1}^n (Y_i - \overline{Y}_n)^2\right\}}}$$

3.1.4 Multivariate version of mean and variance

Let $\boldsymbol{X}_1, \boldsymbol{X}_x, \dots, \boldsymbol{X}_n$ be a random sample from a multivariate distribution F, then the sample mean and sample variance is defined as:

$$\overline{\boldsymbol{X}}_n = \frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i$$
$$\mathcal{S}_n = \frac{1}{n}\sum_{i=1}^n \left(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_n\right)\left(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_n\right)^T$$

$E\left[\overline{\boldsymbol{X}}_n\right] = \frac{1}{n}E\left[\sum_{i=1}^n \boldsymbol{X}_i\right] = E\left[\boldsymbol{X}_i\right]$ for any i.

$$\mathcal{S}_n = \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_n\right)\left(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_n\right)^T$$
$$= \left(\left(\frac{1}{n}\sum_{k=1}^{n}\left(X_{ik} - \overline{X}_i\right)\left(X_{jk} - \overline{X}_j\right)\right)\right)_{ij}$$

For the diagonal element, it resembles sample variance for elements, while the other elements resemble sample covariance.

$$E\left[\frac{n}{n-1}\mathcal{S}_n\right] = E\left[\left(\left(\frac{1}{n-1}\sum_{k=1}^{n}\left(X_{ik} - \overline{X}_i\right)\left(X_{jk} - \overline{X}_j\right)\right)\right)_{ij}\right]$$

For diagonal elements, it becomes the expectation of unbiased sample variance, which is variance σ^2.

Consider for non-diagonal elements, this calculation

$$E\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)\left(Y_i - \overline{Y}_n\right)\right] = E\left[\frac{1}{n}\sum_{i=1}^{n}X_iY_i - \overline{X}_n\overline{Y}_n\right]$$
$$= \frac{1}{n}\sum_{i=1}^{n}E\left[X_iY_i\right] - E[\overline{X}_n\overline{Y}_n]$$
$$= E[XY] - E[\overline{X}_n\overline{Y}_n]$$

$$E[\overline{X}_n\overline{Y}_n] = E\left[\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)\left(\frac{1}{n}\sum_{i=1}^{n}Y_i\right)\right]$$
$$= \frac{1}{n^2}E\left[\sum_{i=1}^{n}\sum_{j=1}^{n}X_iY_j\right]$$
$$= \frac{1}{n^2}\left\{\sum_{i=1}^{n}E\left[X_iY_i\right] + \sum_{i\neq j}X_iY_j\right\}$$
$$= \frac{1}{n^2}\left\{nE[XY] + n(n-1)E[X]E[Y]\right\}$$
$$= \left(\frac{1}{n}\right)E[XY] + \left(\frac{n-1}{n}\right)E[X]E[Y]$$

Why does the last two steps hold? $\longrightarrow$ The samples are i.i.d., and hence $\boldsymbol{X}_i$ is independent from $\boldsymbol{Y}_j$ for $i \neq j$. And hence individual components are also independent.

$$E\left[\frac{n}{n-1}\mathcal{S}_n\right] = \Sigma$$

the variance-covariance matrix.

3.1.5 Sample moments

Let $X_1, \ldots, X_n$ be random sample from a distribution F, then the r^{th} order raw moment and central moment is defined as:

$$m_r' = \frac{1}{n}\sum_{i=1}^{n}X_i^r \quad \text{and} \quad m_r = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^r,\ r > 0$$

$E[m_r'] = E[X_i^r] = \mu_r'$, which is the population r^{th} raw moment.

The sample central moments can be derived from the sample raw moments and vice versa.

3.1.6 Order statistics

Let $X_1, \ldots, X_n$ be a random sample from some distribution F. Then the r-th order statistic $X_{(r)}$ is the r-th smallest element of $X_1, \ldots, X_n$, $r = 1, \ldots, n$

Properties of order statistics

(a) Let $X_i \overset{\text{iid}}{\sim} F$ with pdf f_X and cdf F_X,

$$f_{X_{(1)},\ldots,X_{(n)}}(t_1, \ldots, t_n) = f_{X_1,\ldots,X_n}(t_1, \ldots, t_n) +$$
$$+ f_{X_1,\ldots,X_n}(t_2, t_1, \ldots, t_n) + \cdots +$$
$$+ f_{X_1,\ldots,X_n}(t_n, \ldots, t_1)$$
$$= \left[f_X(t_1)\ldots f_X(t_n)\right]n!$$
$$= n!\prod_{j=1}^{n}f_X(t_j)$$

this holds if $t_1 < t_2 < \cdots < t_n$, otherwise 0.

(b) Let $X_1, \ldots, X_n$ be a random sample from some distribution with CDF F_X. Then the CDF for $X_{(n)}$, is given by

$$F_{X_{(n)}}[t] = F_X^n[t]$$

Proof.

$$F_{X_{(n)}}[t] = P(X_{(n)} \leq t)$$
$$= P(X_1 \leq t, \ldots, X_n \leq t)$$
$$= P(X_1 \leq t) \cdot P(X_2 \leq t) \cdot \cdots \cdot P(X_n \leq t)$$
$$= (P(X_1 \leq t))^n$$
$$= (F_x[t])^n$$

□

(c) Let $X_1, \ldots, X_n$ be a random sample from some distribution with CDF F_X. Then the CDF for $X_{(1)}$, is given by

$$F_{X_{(1)}}[t] = 1 - \{1 - F_X[t]\}^n$$

Proof.

$$\begin{aligned} F_{X_{(1)}}[t] &= P(X_{(1)} < t) \\ &= 1 - P(X_{(1)} > t) \\ &= 1 - P(X_1 > t, X_2 > t, \ldots, X_n > t) \\ &= 1 - \{P(X_1 > t)\}^n \\ &= 1 - \{1 - F_X[t]\}^n \end{aligned}$$

□

(d) Let $X_1, \ldots, X_n$ be a random sample from some distribution with CDF F_X and pdf f_X. Then the pdf of $X_{(r)}$, is given by

$$f_{X_{(r)}}[t] = \frac{n!}{(r-1)!(n-r)!} F_X[t]^{r-1} f_X[t]\{1 - F_X[t]\}^{n-r}$$

A brief proof outline:
$F_{X_{(r)}}[t] = P[X_{(r)} \leq t] = P[\text{atleast } r\text{-many are } \leq t]$ which consists of mutually exclusive events wherein $r,\ r+1\ldots$ samples are $\leq t$, which sums to $\sum_{m=r}^{n} \binom{n}{m} [F_X[t]]^m [1 - F_X[t]]^{n-m}$. Differentiate appropriately to get the desired pdf.

(e) Let $X_1, \ldots, X_n$ be a random sample from some distribution with CDF F_X and pdf f_X. Then the joint pdf of $X_{(r)}$ and $X_{(s)}$ with $r < s$, is given by

$$f_{X_{(r)},X_{(s)}}[w,t] = \begin{cases} 0 & w > t \\ \binom{n}{r-1} [F_X[w]]^{r-1} (n-r+1) f_X[w] \binom{n-r}{s-r-1} [F_X[t] - F_X[w]]^{s-r+1} (n-s+1) f_X[t] [1 - F_X[t]]^{n-s} & \text{o/w} \end{cases}$$

3.1.7 Sample Median

Let $X_1, \ldots, X_n$ be a random sample from some distribution F. Then the sample median $\tilde{X}_{me}$ is given by

$$\tilde{X}_{me} = \begin{cases} X_{((n+1)/2)} & \text{if } n \text{ is odd} \\ \left\{X_{(n/2)} + X_{(n/2+1)}\right\}/2 & \text{if } n \text{ is even} \end{cases}$$

3.2 Large Sample Results

3.2.1 Weak Law of Large Numbers, WLLN

Let $X_1, \ldots, X_n$ be a random sample from a population with $E[g(X)] = \eta < \infty$, where $g : \mathbb{R} \to \mathbb{R}$ is a function. Then $\frac{1}{n}\sum_{i=1}^{n} g(X_i) \xrightarrow{p} \eta$ as $n \to \infty$, i.e., for any $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} g(X_i) - \eta\right| > \epsilon\right) = 0$$

3.2.2 Central Limit Theorem, CLT

Let $X_1, \ldots, X_n$ be a random sample from some distribution with expectation μ and finite variance σ^2. Then $\mathcal{T}_n^{\mu,\sigma} = \sqrt{n}(\overline{X}_n - \mu)/\sigma \xrightarrow{d} \mathcal{N}(0,1)$, that if for any $x \in \mathbb{R}$, the CDF of $\mathcal{T}_n^{\mu,\sigma}$, say G_n satisfies

$$G_n(x) \to \Phi(x), \quad \text{as} \quad n \to \infty$$

where Φ is the CDF of $\mathcal{N}(0,1)$ distribution.