








MAS-1

Study Review

Nicholas Langevin

18 mars 2019

-  Probability Review
-  Stochastic Processes
-  Life Contingencies
-  Simulation
-  Statistics
-  Extended Linear Model
-  Time Series

Lesson 1 : Probability Review

- > **Bernoulli Shortcut** : If a random variable can only assume two values a and b with probability q and $1 - q$, then its variance is $q(1 - q)(b - a)^2$

Lesson 2 : Parametric Distributions

> Transformations :

- Transformed : $\tau > 0$
- Inverse : $\tau = -1$
- Inverse-Transformed : $\tau < 0, \tau \neq 1$

Lesson 4 : Markov Chains

> Chapman-Kolmogorov :

$$P_{ij}^{(n+m)} = \sum_{k=0}^{\infty} P_{ik}^{(n)} P_{kj}^{(m)}$$

> Gambler's ruin :

$$p_j = \begin{cases} \frac{j}{N} & , r = 1 \\ \frac{r^j - 1}{r^N - 1} & , r \neq 1 \end{cases}$$

où $r = \frac{q}{p}$, p : winning prob.

- > **Algorithmic efficiency** : with N_j = number of steps from j^{th} solution to best solution.

$$E[N_j] = \sum_{i=1}^{j-1} \frac{1}{i}$$

$$\text{Var}(N_j) = \sum_{i=1}^{j-1} \left(\frac{1}{i} \right) \left(1 - \frac{1}{i} \right)$$

As $j \rightarrow \infty$, $E[N_j] \rightarrow \ln j$, $\text{Var}(N_j) \rightarrow \ln j$

Lesson 5 : Markov Chain Classification

- > An **absorbing** state is one that cannot be exited.
- > State j is **accessible** ($i \rightarrow j$) from state i if $p_{ij}^n > 0$, $\forall n \geq 0$.
- > Two states **communicate** if $i \leftrightarrow j$.
- > A **class** of states is a maximal set of state that communicate with each other.
- > A Markov chain is **irreducible** if it has only one class.
- > A state (class) is **recurrent** if the probability of reentering the state is 1. $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$
- > A state (class) is **transient** if it is not recurrent. $\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$
- > A finite Markov Chain must have at least one recurrent class. If it is irreducible, then it is recurrent.

Lesson 6 : Markov Chains Limiting Probability

- > A chain is **positive recurrent** is the expected number of transitions until the state occur is finite, **null recurrent** otherwise. Null recurrent mean that the long-term proportion of time in each state is 0.
- > A chain is **periodic** when states occur every n periods for $n > 1$.
- > A chain is **aperiodic** when the period is 1. In other world, $P_{ii}^{(1)} > 0, \forall i$
- > A chain is **ergodic** when the chain is aperiodic and positive irreducible recurrent.
- > **Stationary probability** :

$$\pi_j = \sum_{i=1}^n P_{ij} \pi_i \quad \sum_{i=1}^n \pi_i = 1$$

- > **Limiting probabilities** : if the chain is ergodic, then

$$P^{(\infty)} = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 \\ \pi_1 & \pi_2 & \pi_3 \\ \pi_1 & \pi_2 & \pi_3 \end{pmatrix}$$

Lesson 7 : Time in Transient States

- > Tips : **Inverting a matrix**
- > $S = (I - P_{\text{transient}})^{-1}$, where s_{ij} is the time in state j given that the current state is i .
- > $f_{ij} = \frac{s_{ij} - \delta_{i,j}}{s_{jj}} = \sum_{n=1}^{\infty} f_{ij}^{(n)}$, where f_{ij} is the probability that state i ever transitions to state j .

Lesson 8 : Branching Processes

- > A branching process is a special type of Markov chain representing the growth or extinction of a population.
- > $E[X_n] = E[Z]^n$, where $E[Z]$ is the expected number of people born in a generation.
- > $\text{Var}(X_n) = \text{Var}(Z) \cdot E[Z]^{n-1} \sum_{k=1}^n E[Z]^{k-1}$
- > If $X_0 \neq 1$ mean and variance of X_n need to be multiplied by X_0 .

> Probability of extinction :

$$\pi_0 = \sum_{j=1}^{\infty} p_j \pi_0^j$$

$$- \mu \leq 1 \Rightarrow \pi_0 \geq 1, \text{ if } X_0 = 1.$$

$$- \mu > 1 \Rightarrow \pi_0 < 1, \text{ if } X_0 = 1.$$

For cubic equation, it guaranteed to factor ($\pi_0 - 1$). Tips : **Synthetic Division**

Lesson 9 : Time Reversible

- > If Q is the reverse-time Markov chain for ergodic P , then $\pi_i Q_{ij} = \pi_j P_{ji}$ with $P_{ii} = Q_{ii}$ and if $p_{ij} = 0 \Leftrightarrow q_{ji} = 0$
- > If $Q = P$, then P is said to be **time-reversible**.

Lesson 10 : Exponential Distribution

> Lack of memory :

$$\Pr(X > k + x | X > k) = \Pr(X > x)$$

- > **Minimum** : if $X_i \sim \text{Exp}(\lambda_i)$, then

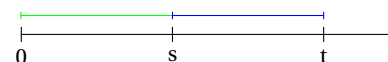
$$\min(X_1, X_2, \dots, X_n) \sim \text{Exp}\left(\sum_{i=1}^n \lambda_i\right)$$

- > The sum of 2 Exponentials random variables is the sum of the maximum and the minimum, since one must be the min and the other the max.

$$X_1 + X_2 = \min(X_1, X_2) + \max(X_1, X_2)$$

Lesson 11 : Poisson Process

- > $X(t) \sim \text{Poisson}[m(t)]$, where $m(t)$ is **mean value function** representing the mean of the number events before time t .
- > Poisson process can't decrease over time. $N(t) \geq N(s)$
- > $N(0) = 0$
- > Increment are **independent** :



$$\Pr[N(t) - N(s) = n | N(s) = k] = \Pr[N(t) - N(s) = n]$$

> Non-homogeneous Poisson process :

$$m(t) = \int_0^t \lambda(u) du$$

where $\lambda(t)$ is the **intensity function**

- > **Homogeneous Poisson process** : The Poisson process is said to be homogeneous when the intensity function is a constant.

$$m(t) = \int_0^t \lambda du = \lambda t$$

We then say that the process have **stationary increments**.

$$\Pr[N(s)] = \Pr[N(t) - N(s)]$$

Lesson 12 : Poisson Process Time To Next Events

- > T_n is the time between the n^e event and the $(n-1)^e$ event.
- > $S_n = \sum_{i=1}^n T_i$, is the time for the n^e event.
- > $F_{T_1}(t) = 1 - e^{-\int_0^t \lambda(u) du}$
- > For homogeneous process : $T_n \sim \text{Exp}(\lambda)$
 $S_n \sim \text{Gamma}(n, \lambda)$

Lesson 13 : Poisson Process Counting Special Type

- > If event of type 1 occur with probability $\alpha_1(t)$, then the event follow a Poisson process with intensity $\lambda(t) \cdot \alpha_1(t)$.

$$m(t) = \int_0^t \lambda(u) \alpha_1(u) du$$

Lesson 14 : Poisson Process Other Characteristics

- > Only for homogeneous Poisson processes.
- > The probability of k event from process 1 is given by :

$$k \sim \text{Binomial}\left(k+l-1, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right)$$

Then the probability that k event from process 1 occur before l from process 2 is :

$$\sum_{i=k}^{k+l-1} \binom{k+l-1}{i} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^i \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{k+l-1-i}$$

- > Given that exactly $N(t) = k$ Poisson events occurred before time t , the joint distribution of event time is the joint distribution of k independent uniform random variables on $(0, t)$.

$$F_{S_1, \dots, S_n | n(t)}(s_1, \dots, s_n | k) = \frac{k!}{t^k}$$

- > For k independent uniform random variable on $(0, t)$, the expected value of the j^{th} order statistics is : $E[T^{(j)}] = \frac{jt}{(k+1)}$.
- > Tips : **Statistic Order**

Lesson 15 : Poisson Process Sums and Mixtures

- > A **Sums** of independent Poisson random variables is a Poisson random with intensify function $\lambda(t) = \sum \lambda_i(t)$. **Warning : Subtraction don't give a Poisson random variable.**
- > A **Mixture** of Poisson processes is not a Poisson processes.
 - **Discrete** mixture :

$$F_{X(t)}(t) = \sum_i w_i F_{X_i(t)}(t)$$
 where $w_i > 0, \sum w_i = 1$
 - **Continuous** mixture :

$$F_{X(t)}(t) = \int F_{X_u(t)}(t) f(u) du$$
 - If $N(t)|\lambda$ is a Poisson random variable and $\lambda \sim \text{Gamma}(\alpha, \theta)$, then $N(t) \sim \text{NegBin}(r = \alpha, \beta = \theta t)$.

Lesson 16 : Compound Poisson Processes

- > A **compound** random variable S is define by $S = \sum_{i=1}^N X_i$ where N is the **primary** distribution and X the **secondary** distribution.

- > If $N(t)$ is a Poisson process, then $S(t)$ is a compound Poisson process with :

$$\begin{aligned} - E[S(t)] &= \lambda t E[X] \\ - \text{Var}(S(t)) &= \lambda t E[X^2] \end{aligned}$$

- > If X_i is discrete, we can separate the process into a sum of subprocess view in **Lesson 13 : Poisson Process Counting Special Type**.

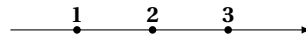
- > **Sums of compound** homogeneous Poisson process is also a Poisson process with :

$$\begin{aligned} - N(t) &\sim \text{Pois}(\sum \lambda_i) \\ - F_X(x) &= \sum_i w_i F_{X_i(t)}(t), \quad w_i = \frac{\lambda_i}{\sum \lambda_i} \end{aligned}$$

Lesson 17 : Reliability Structure Functions

- > $\phi(\mathbf{x})$ is the **structure** function for a system. It equal 1 if the systeme function, 0 otherwise.

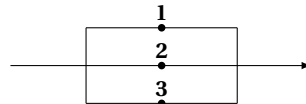
- > A **series** system is define as a **minimal path set**. The system is working if all components are working.



The serie structure function is define as

$$\phi(\mathbf{x}) = \prod_{i=1}^n x_i$$

- > A **parallel** system is define as a **minimal cut set**. The systeme is working if at least 1 components is working.



The parallel structure function is define as

$$\phi(\mathbf{x}) = 1 - \prod_{i=1}^n (1 - x_i)$$

- > Tips : Minimal path set is all way for the system to work, and the minimal cut set is all the way for the system to not work.
- > Tips : If set is $\{1, 2, 3\}$ and $\{1, 2\}$, the *minimal* mean we only take $\{1, 2\}$.
- > Tips : *Minimal cut* is a serie of parallel structure and *minimal path* is a parallel of serie structure.

Lesson 18 : Reliability Probabilities

- > $r(\mathbf{p})$ is the same polynomial as $\phi(\mathbf{x})$.
- > Inclusion/exclusion bounds using minimal path :

$$\begin{aligned} r(\mathbf{p}) &\leq \sum A_i \\ r(\mathbf{p}) &\geq \sum A_i - \sum A_i \cup A_j \end{aligned}$$

$$r(\mathbf{p}) \leq \sum A_i - \sum A_i \cup A_j + \sum A_i \cup A_j \cup A_k$$

where $A_i = \sum p_i$ is the probability of the i^{e} minimal path set work.

- > Inclusion/exclusion bounds using minimal cut :

$$1 - r(\mathbf{p}) \leq \sum A_i$$

$$1 - r(\mathbf{p}) \geq \sum A_i - \sum A_i \cup A_j$$

$$1 - r(\mathbf{p}) \leq \sum A_i - \sum A_i \cup A_j + \sum A_i \cup A_j \cup A_k$$

where $A_i = \sum (1 - p_i)$ is the probability of the i^{e} minimal cut set work.

- > Bounds using intersections :

$$\prod \phi(\mathbf{X})^{\text{min. cut}} \leq r(\mathbf{p}) \leq \prod \phi(\mathbf{X})^{\text{min. path}}$$

- > **Random graph** :

$$1 - P_n = \sum_{k=1}^{n-1} \binom{n-1}{k-1} q^{k(n-k)} p_k$$

$$1 - P_n \leq (n+1) q^{n-1}$$

$$P_1 = 1$$

Lesson 19 : Reliability Time to Failure

- > Expected amount of time to failure :

$$E[\text{system life}] = \int_0^\infty r(\bar{F}(t)) dt$$

where,

- For serie system :

$$r(\bar{F}(t)) = \prod_{i=1}^n \bar{F}_i(t)$$

- For parallel system :

$$r(\bar{F}(t)) = 1 - \prod_{i=1}^n F_i(t)$$

- > **Shortcut** : k out of n system with exponentials(θ) : $E[T] = \theta \sum_{i=k}^n \frac{1}{i}$

- > **Hazard rate function** (failure rate function) :

$$h(t) = \frac{f(t)}{\bar{F}(t)}$$

and we say that the distribution

- is an increasing failure rate if $h(t)$ is non-decreasing function of t .
- is an decreasing failure rate if $h(t)$ is non-increasing function of t .

- > **Cumulative hazard function** :

$$H(t) = \int_0^t h(u) du = -\ln \bar{F}(t)$$

with $\frac{H(t)}{t}$ the average of the hazard rate.

Lesson 20 : Survival Models

$${}_t p_x = \frac{\ell_{x+t}}{\ell_x}, \quad {}_t q_x = \frac{\ell_x - \ell_{x+t}}{\ell_x}$$

$${}_t |u q_x = \frac{\ell_{x+t} - \ell_{x+t+u}}{\ell_x}$$

$${}_t + u p_x = {}_t p_x \cdot {}_t p_{x+u}$$

$${}_t |u q_x = {}_t + u q_x - {}_t q_x = {}_t p_x \cdot {}_t + u q_{x+t}$$

- > Let be N_x the number of life surviving to age x , then

$$(N_{x+t} | N_x = n) \sim \text{Bin}(n, {}_t p_x)$$

- > **Force of mortality** :

$$\mu_{x+t} = \frac{f_{T_x}(t)}{{}_t p_x} = -\frac{d}{dt} \ln {}_t p_x$$

Linear interpolation(D.U.D) :

$$\ell_{x+t} = (1-t)\ell_x + t\ell_{x+1}$$

Shortcut : $\forall t \in (0, 1), \forall x \in \mathbb{N}, x < x+t < x+1 :$

$$\rightarrow tq_x = t \cdot q_x$$

$$\rightarrow \mu_{x+t} = \frac{q_x}{1-t \cdot q_x}$$

- > **Expected life time** : Let $k_x = \lfloor T_x \rfloor$, the full years until death. Then e_x is the **curtate life expectancy** and \bar{e}_x the **complete life expectancy**. ω is the age where $\ell_\omega = 0$ and $\omega = \infty$ by convention is nothing is said.

$$e_x = E[K_x] = \sum_{k=1}^{\omega-x-1} k p_x$$

$$\bar{e}_x = E[T_x] = \int_0^{\omega-x} t p_x dt \stackrel{\text{D.U.D}}{=} e_x + 0.5$$

Lesson 21 : Contingent Payments

The contract here are define with K_x to pay at the end of death year. All same contract can be define with T_x to pay at the moment of death. Then we use integral instead of sum and use

$$\Pr(K = k) = k p_x q_{x+k} \Rightarrow f_{T_x}(t) = t p_x \mu_{x+t}$$

Life Insurance :

- Whole Life insurance :

$$A_x = \sum_{k=0}^{\infty} v^{k+1} k p_x q_{x+k}$$

- Term Life insurance :

$$A_{x:\overline{n}|} = \sum_{k=0}^n v^{k+1} k p_x q_{x+k}$$

- Deferred insurance :

$${}_m|A_x = \sum_{k=m}^{\infty} v^{k+1} k p_x q_{x+k}$$

- Endowment insurance :

$$A_{x:\overline{n}|} = A_{x:\overline{n}|}^1 + n E_x$$

- Pure Endowment :

$${}_n E_x = v^n n p_x$$

Life Annuities :

- Whole Life annuity

$$\ddot{a}_x = \sum_{k=0}^{\infty} v^k k p_x$$

- Temporary Life annuity

$$\ddot{a}_{x:\overline{n}|} = \sum_{k=0}^n v^k k p_x$$

- Deferred annuity

$${}_m|\ddot{a}_x = \sum_{k=m}^{\infty} v^k k p_x$$

- Certain and life annuity

$$\ddot{a}_{x:\overline{n}|} = \ddot{a}_{\overline{n}|} + {}_m|\ddot{a}_x$$

Illustrative Life Table :

- $A_x = v^n q_x + p_x A_{x+1}$
- $\ddot{a}_x = 1 + v p_x \ddot{a}_{x+1}$
- $A_{x:\overline{n}|} = A_x - n E_x A_{x+n}$
- $\ddot{a}_{x:\overline{n}|} = \ddot{a}_x - n E_x \ddot{a}_{x+n}$
- ${}_m|A_x = m E_x A_{x+m}$
- ${}_m|\ddot{a}_x = m E_x \ddot{a}_{x+m}$
- $\ddot{a}_x = 1 + a_x$
- $A_x = 1 - d \ddot{a}_x$

- > **Joint life annuity**(\ddot{a}_{xy}) make payments until the earliest death pf two lives.

- > **Last survivor annuity**($\ddot{a}_{\overline{xy}}$) make payments until the last death of two lives.

$$\ddot{a}_x + \ddot{a}_y = \ddot{a}_{xy} + \ddot{a}_{\overline{xy}}$$

- > **Premiums :**

$$M \cdot A_x = P \ddot{a}_x$$

$$P = \frac{M \cdot A_x}{\ddot{a}_x} = \frac{M}{\ddot{a}_x} - M \cdot d$$

Lesson 22 : Simulation Inverse Method

- > **Linear congruential generators :**

$$x_k = (a x_{k-1} + c) \bmod m$$

$$x_k = b - \left\lfloor \frac{b}{m} \right\rfloor m$$

where $b = (a x_{i-k} + c)$ and $x_0 \equiv \text{seed}$

- > **Inverse transformation method :**

$$\Pr(F^{-1}(u) \leq x) = \Pr(u \leq F(x)) = F(x)$$

then $x = F^{-1}(u)$ where $U \sim \text{Unif}(0, 1)$

- Normal Case : $x = \mu + \sigma z$

- Log-Normal Case : $x = e^{\mu + \sigma z}$

where $z = \Phi^{-1}(u)$, with linear interpolation.

- > **Tips : Discrete Cumulative Function**

- > **Tips :** if $\uparrow U \equiv \downarrow X$ then $(1 - u_i) \Rightarrow u_i$

Lesson 23 : Simulation Application

- > $\Pr(X \leq x) \approx \frac{1}{m} \sum_{j=1}^m \mathbb{1}_{\{x^{(j)} \leq x\}}$

- > $E[X^k] \approx \frac{1}{m} \sum_{j=1}^m [x^{(j)}]^k$

- > $\text{VaR}_k(X) \approx X^{[j]}$

- > $\text{TVaR}_k(X) \approx \frac{1}{m(1-k)} \sum_{j=j_0+1}^m X^{(j)} \mathbb{1}_{\{X^{(j)} > X^{[j_0]}\}}$
 $\approx \frac{1}{m-j_0} \sum_{j=j_0+1}^m X^{[j]}$

where

- $j_0 = \lfloor m \cdot k \rfloor$
- m is the number of simulations.
- $X^{(j)}$ is the j^{th} simulations.
- $X^{[j]}$ is the j^{th} simulations in order statistics.

Lesson 24 : Simulation Rejection Method

- > **General method** : Let $f(x)$ be the density function of variable to simulate, and let $g(x)$ be the **base distribution**, a random density function that is easy-to-simulate with nonzero wherever $f(x) \neq 0$.

$$c = \max \frac{f(x)}{g(x)}$$

Generate two uniform number u_1, u_2 . Let

$x = G^{-1}(u_1)$. Accept x_1 only if

$$u_2 \leq \frac{f(x_1)}{c \cdot g(x_1)}$$

- > **Simulating gamma distribution** : Use $\text{Exp}(\alpha \cdot \theta)$ as the base distribution and $x = \alpha \cdot \theta$ that maximize c .

- > **Simulating standard normal distribution** : Generate 3 uniform u_1, u_2, u_3 . Let $y_1 = -\ln u_2$ and $y_2 = -\ln u_3$. Accept y_1 if

$$y_2 \geq \frac{(y_1 - 1)^2}{2}$$

and add $(-)$ if $u_3 \geq 0.5$

- > The **Number of iteration** is a Ross-geometric distribution with mean c . Let be β the mean of a geometric distribution given in the exam appendix :

$$E[N] = 1 + \beta = c$$

$$\text{Var}(N) = \beta(1 + \beta)$$

Lesson 25 : Estimator Quality

- > **Bias** : This quality measures if, on average, the estimator is on the expected value of the parameter.

$$E[\hat{\theta}] = \theta + \text{bias}_{\hat{\theta}}(\theta)$$

- If $\text{bias}_{\hat{\theta}}(\theta) = 0$, then $\hat{\theta}$ is **unbiased**.

- If $\lim_{n \rightarrow \infty} \text{bias}_{\hat{\theta}}(\theta) = 0$, then $\hat{\theta}$ is **asymptotically unbiased**.

- If $\text{bias}_{\hat{\theta}}(\theta) \neq 0$, then $\hat{\theta}$ is **biased**.

- > **Consistency** : This quality measures if the probability that the estimator is different from the parameter by more than ε goes to 0 as n goes to infinity.

$$\lim_{n \rightarrow \infty} \Pr(|\hat{\theta} - \theta| > \varepsilon) \rightarrow 0, \forall \varepsilon > 0$$

In other word, as $n \rightarrow \infty$, $E[\hat{\theta}] \rightarrow \theta$, $\text{Var}(\hat{\theta}) \rightarrow 0$

- > **Efficiency** : This quality measures the variance of the estimator. Lower the variance is, more efficient is the estimator.

$$\text{Efficiency of } \hat{\theta} = \frac{\text{Var}(\hat{\theta})^{\text{rao}}}{\text{Var}(\hat{\theta})}$$

$$\text{Relative efficiency of } \hat{\theta}_1 \text{ to } \hat{\theta}_2 = \frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\hat{\theta}_1)}$$

See the **rao-cramer lower bound**.

- > **Mean Square Error** : This quality measures the expected value of the square difference between the estimator and the parameter.

$$\text{MSE}_{\hat{\theta}}(\theta) = E[(\hat{\theta} - \theta)^2] = (\text{bias}_{\hat{\theta}}(\theta))^2 + \text{Var}(\hat{\theta})$$

- > An estimator is called a **uniformly minimum variance unbiased estimator(UMVUE)** if it's unbiased and if there is no other unbiased estimator with a smaller variance for any true value θ .

- > Some estimator :

- $\bar{x} = \frac{1}{n} \sum x_i$ is a unbiased estimator of the mean μ . $\text{Var}(\bar{x}) = \frac{1}{n} \text{Var}(x)$

- $s^2 = \sum \frac{(x_i - \bar{x})^2}{n-1}$ is a unbiased estimator of the variance σ^2 .
- $\hat{\sigma}^2 = \sum \frac{(x_i - \bar{x})^2}{n}$ is an asymptotically unbiased of the variance σ^2 .
- $\hat{\mu}'_k = \frac{1}{n} \sum x_i^k$, where $\hat{\mu}'_1 = \bar{x}$ and $\hat{\mu}_k = \frac{1}{n} \sum (x_i - \bar{x})^k$, where $\hat{\mu}_1 = 0$ and $\hat{\mu}_2 = \hat{\sigma}^2$.

Lesson 26 : Kernel Density Estimation

- > **Empirical distribution** : All data is assigning a probability of $\frac{1}{n}$. This is the same method used for simulation, see **Lesson 23 : Simulation Application**.

$$F_e(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{x_i \leq x\}}$$

$$f_e(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{x_i = x\}} \\ = F_e(x) - F_e(x_{i-1})$$

- > **Kernel Density** is a empirical distribution smoothed with a base function. Let define the scaling factor b called **bandwidth**.

- The kernel-density estimate of the density function is : $\hat{f}(x) = \frac{1}{n} \sum k\left(\frac{x-x_i}{b}\right) \Leftrightarrow \sum f_e(x) k\left(\frac{x-x_i}{b}\right)$
- The kernel-density estimate of the distribution is : $\hat{F}(x) = \frac{1}{n} \sum K\left(\frac{x-x_i}{b}\right)$

- > **Rectangular(uniform, box) kernel** :

$$k(x) = \begin{cases} \frac{1}{2b}, & -1 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

$$K(x) = \begin{cases} 0, & x < -1 \\ 0.5(x+1), & -1 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$$

$$\hat{f}(x) = \frac{F_e(x+b) - F_e(x-b)}{2b}$$

- > **Triangular kernel** :

$$k(x) = \begin{cases} \frac{1-|x|}{b}, & -1 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

$$K(x) = \begin{cases} 0, & x < -1 \\ \frac{(1+x)^2}{2}, & -1 \leq x \leq 0 \\ 1 - \frac{(1-x)^2}{2}, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$$

- > **Gaussian kernel** : The distribution become normal with $\mu = x_i$ and $\sigma = b$.

$$k(x) = \frac{e^{-x^2/2}}{b\sqrt{2\pi}}$$

$$K(x) = \Phi(x)$$

- > Other kernel : $k(x) = \beta(x)$ and $K(x) = B(x)$

- > **kernel moments** : Let X be the kernel density estimate and x_i the empirical estimate.

We then condition on x_i .

$$E[X] = E[E[X|x_i]] = E[x_i]$$

$$\text{Var}(X_R) = \text{Var}(x_i) + \frac{b^2}{3}$$

$$\text{Var}(X_T) = \text{Var}(x_i) + \frac{b^2}{6}$$

$$\text{Var}(X_G) = \text{Var}(x_i) + b^2$$

- > Tips : For rectangular kernel, $E[x|x_i]$ is a uniform($x_i - b, x_i + b$).

Lesson 27 : Method of Moments

- > **Types of data** :

- Complete data : Data is complete if we are given the exact value of each observation.
- Grouped data : Set of interval and we know how many observation are in each.
- Censored data : Value that are in a interval, but we don't know the exact value. Like limits ($\min(X, u)$).
- Truncated data : We have data only when it in certain range, otherwise we don't know. Like deductible ($X|X > d$).

- > **Method of Moments** : We match $\hat{\mu}'_k = E[X^k]$ and find the parameters. If data is Censored or Truncated, we need to match the censored or truncated moment : $\hat{\mu}'_k = E[\min(X, u)^k]$ or $\hat{\mu}'_k = E[X^k | X > d]$.

- > For pareto distribution, if $\hat{\mu}'_2 = \hat{\sigma}^2 + \bar{x}^2 \leq 2\bar{x}^2$, the method of moment is unstable and can't be used.

Lesson 28 : Percentile Matching

- > **Percentile Matching** : We match $F_e(\hat{\pi}_p) = p$ and find the parameters.

- For censored data, we need select percentiles within the range of the uncensored portion of the data.
- For truncated data, we need to match the percentiles of the conditional distribution :

$$F(x|X > d) = \frac{\Pr(d < X \leq x)}{\Pr(X > d)} = \frac{F(x) - F(d)}{1 - F(d)}$$

$$S(x|X > d) = \frac{S(x)}{S(d)}$$

- > **Smoothed empirical percentile** :

$$\hat{\pi}_p = (1-h)X^{[j]} + hX^{[j+1]}$$

where

- $j = \lfloor (n+1)p \rfloor$
- $h = (n+1)p - j$
- $X^{[j]}$ is the j^{th} order statistics.

Lesson 29 : Maximum Likelihood Estimators

- > **Maximum Likelihood Estimators** : We maximize the probability of observing the data.

$$L(\theta) = \prod g(x_i; \theta)$$

$$l(\theta) = \sum \ln g(x_i; \theta)$$

- Individual data : $g(x_i; \theta) = f(x_i)$
- Grouped data : $g(x_i; \theta) = F(x_i) - F(x_{i-1})$
- Censored data : $g(x_i; \theta) = S(x_i)$
- Truncated data : $g(x_i; \theta) = \frac{f(x)}{s(x)}$

Lesson 30 : MLE Special Techniques

- > Case MLE equals MME

- For Exponential, $\hat{\theta}^{\text{MLE}} = \bar{x}$
- For Gamma with fixed α , $\hat{\theta}^{\text{MLE}} = \hat{\theta}^{\text{MME}}$
- For Normal, $\hat{\mu}^{\text{MLE}} = \bar{x}$ and $(\hat{\sigma}^2)^{\text{MLE}} = \frac{1}{n} \sum (x_i - \bar{x})^2$
- For Binomial, $m q = \bar{x}$ then given m , $\hat{q}^{\text{MLE}} = \frac{\bar{x}}{m}$
- For Poisson, $\hat{\lambda}^{\text{MLE}} = \hat{\lambda}^{\text{MME}}$
- For Binomial Negative, given r or β , $(r\beta)^{\text{MLE}} = \bar{x}$

- > Parametrization and Shifting :

- Parametrization : MLE's are independent of parametrization $\lambda = \frac{1}{\theta} \Leftrightarrow \hat{\lambda}^{\text{MLE}} = \frac{1}{\hat{\theta}^{\text{MLE}}}$

- Shifting the distribution is equivalent of shifting the data.

- > Transformations : MLE's are invariant under one-to-one transformation. Then if we have a transformed variable, we can untransform the data and find the parameter of the untransform distribution.

Tips : **Transformations of distribution**

- > Weibull distribution : If the data is censored(u) or truncated(d), then

$$\left(\hat{\theta}^{\text{MLE}}\right)^{\tau} = \frac{\sum (x_i - d_i)^{\tau}}{\sum \mathbb{1}_{\{x_i \leq u\}}}$$

if $\tau = 1$, then the distribution is Exponential.

- > Pareto distribution with fixed $\theta : \hat{\alpha} = \frac{n}{K}$

$$K = \sum_{i=1}^{n+c} \ln(\theta + d_i) - \sum_{i=1}^{n+c} \ln(\theta + x_i)$$

where $n \equiv$ number of non-censored(c) data.

- > Single-parameter Pareto : $\hat{\alpha} = \frac{n}{K}$

$$K = \sum_{i=1}^{n+c} \ln \max(\theta, d_i) - \sum_{i=1}^{n+c} \ln x_i$$

where $n \equiv$ number of non-censored(c) data.

- > Uniform(0, θ) : We take the smallest θ possible, $\hat{\theta}^{\text{MLE}} = \max(x_1, \dots, x_n)$

- Censored(u) : $\hat{\theta}^{\text{MLE}} = \frac{nd}{\sum \mathbb{1}_{\{x_i < d\}}}$

- Grouped : We take the highest interval (L, U). $\hat{\theta}^{\text{MLE}} = \min(U, \hat{\theta}_{\text{Censored(L)}}^{\text{MLE}})$
- > Bernoulli : Let have a random variable that can take 2 values, n and m . Then $\hat{p} = \frac{n}{n+m}$
- > Tips : If $L(\theta)$ look like a density distribution, $\hat{\theta}^{\text{MLE}} \equiv \text{mode of this distribution}$.

Lesson 31 : Variance of MLE

- > **Fisher information matrix :**

$$I(\theta) = nE \left[\left(\frac{d \ln f(x; \theta)}{d\theta} \right)^2 \right]$$

$$= -nE \left[\frac{d^2 \ln f(x; \theta)}{d\theta^2} \right]$$

using the loglikelihood function

$$I(\theta) = E \left[\left(\frac{d l(x_1, \dots, x_n; \theta)}{d\theta} \right)^2 \right]$$

$$= -E \left[\frac{d^2 l(x_1, \dots, x_n; \theta)}{d\theta^2} \right]$$

- > **Rao-Cramer lower bound** is the lowest possible variance for a unbiased estimator $\hat{\theta}$ of θ . Then $\hat{\theta} \sim \text{Normal}(0, \text{Var}(\hat{\theta})^{\text{rao}})$

$$\text{Var}(\hat{\theta})^{\text{rao}} \geq \frac{1}{I(\theta)}$$

under certain regularity conditions

- The second derivative of the loglikelihood exist.
- The support of the density function is not function of θ .

Lesson 32 : Sufficient Statistics

- > A **sufficient statistics** are statistics that capture all the information about the parameter we are estimating that the sample as to offer.
- > A statistics is sufficient when the distribution of a sample given a statistics does not depend on the parameter. Y is a sufficient statistics for a parameter θ if and only if

$$L(x_1, \dots, x_n; \theta | Y) = h(x_1, \dots, x_n)$$

$$L(x_1, \dots, x_n; \theta) = g(y)h(x_1, \dots, x_n)$$

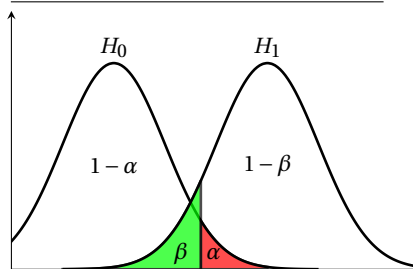
where $h(x_1, \dots, x_n)$ is a function that does not involve θ .

- > **Rao-Blackwell Theorem :** For any unbiased estimator $\hat{\theta}$ and sufficient statistic Y , the estimator $E[\hat{\theta} | Y]$ is unbiased and has variance less than or equal to $\text{Var}(\hat{\theta})$.
- > The maximum likelihood estimator is a function of a sufficient statistic.

Lesson 33 : Hypothesis Testing

- > Let be H_0 the **null hypothesis** and H_1 the **alternative hypothesis**.

	Accept H_0	Reject H_0
H_0 True	$1 - \alpha$	α
H_1 True	β	$1 - \beta$



- > The α value is usually name :

- Probability of Type I error
- Size of critical region
- significance level

The β value is usually name :

- Probability of Type II error

The $(1 - \beta)$ value is usually name :

- The power of test.

- > We will reject H_0 in favor of H_1 if a certain condition occurred ($X > \gamma$), named the **critical region**. Then the probability of rejecting H_0 is giving by

$$\Pr(X > \gamma | H_0 \equiv \text{true}) = \alpha$$

- > Lowering the probability of type I error came at the cost of raising the probability of type II error. One way to lower both is to increase sample size.

- > The **p-value** is the probability of the observation if H_0 is true. H_0 is rejected if and only if the p-value is less than the significance level.

$$P_{\text{value}} < \alpha$$

Lesson 34 : Confidence Interval and Sample Size

- > Let be c the **confidence coefficient**. Then we can say we're 100% confident that the parameter is between (a, b) , called the **confidence interval**. $\alpha = 1 - c$

$$\theta \in \hat{\theta} \pm z_{1+\frac{c}{2}} \sqrt{\text{Var}(\hat{\theta})}$$

- > We can find the probability that the half-width of the interval is less than k .

$$\Pr(|\hat{\theta} - \theta| \leq k) \geq \frac{1+c}{2}$$

$$\Phi \left(\frac{k}{\sqrt{\sigma^2/n}} \right) \geq \frac{1+c}{2}$$

- > To find the sample size needed to have a certain (α) and $(1 - \beta)$, we resolve

$$\Pr(\bar{x} > k | H_0) = 1 - \Phi \left(\frac{k - \mu_0}{\sqrt{\sigma^2/n}} \right) = \alpha$$

$$\Pr(\bar{x} > k | H_1) = 1 - \Phi \left(\frac{k - \mu_1}{\sqrt{\sigma^2/n}} \right) = 1 - \beta$$

Lesson 35 : Confidence Intervals for Means

- > The **chi-square** is a one-parameter family distribution. In definition, it is a gamma with $\alpha = \frac{n}{2}$ and $\theta = 2$. The only parameter n is called **degree of freedom**.

- Let $X_i, i = 1, \dots, n$ be normal random variable with mean μ and variance σ^2 .

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

- Let $x_i, i = 1, \dots, n, n \geq 2$ be random sample from normal distribution with variance σ^2 .

$$W = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

- > The **student** is a one-parameter family distribution. We define it as

$$T_{(n)} = \frac{Z}{\sqrt{W/n}}$$

where $Z \sim N(0, 1)$ and $W \sim \chi^2_{(n)}$.

Note that as $n \rightarrow \infty, T \rightarrow N(0, 1)$

- > When the variance is unknown, we need to estimate it with the unbiased estimator S^2 .

$$T_{(n-1)} = \frac{\bar{x} - \mu}{\sqrt{S^2/n}}$$

- > Testing the difference of means from two populations.

- Two independent normal distributions :

$$x_1, \dots, x_n \sim N(\mu_x, \sigma_x^2)$$

$$y_1, \dots, y_m \sim N(\mu_y, \sigma_y^2)$$

$$T_{(n+m-2)} = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{S \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

$$\text{where } S^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{m+n-2}$$

- Two dependent distributions :

$$(x_1, y_1), \dots, (x_n, y_n) \sim \text{bivariate Normal dis}$$

Appendix

Inverting a matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Ajouter pour une matrice 3x3

Synthetic Division

Exemple : Factorize $x^3 - 12x^2 - 81$

$$\begin{array}{r|rrrr} 3 & 1 & -12 & 0 & -81 \\ & & 3 & -27 & -81 \\ \hline & 1 & -9 & -27 & 0 \end{array}$$

then, $x^3 - 12x^2 - 81 = (x-3)(x^2 - 9x - 27)$

Deductible and Limite

$$\begin{aligned} X &= \min(X; d) + \max(0; X - d) \\ E[X] &= E[\min(X; d)] + E[\max(0; X - d)] \\ &= E[(X \wedge d)] + E[(x - d)_+] \\ &= E[(X \wedge d)] + e_x(d) \cdot S_x(d) \end{aligned}$$

Statistic Order

- $Y_1 = \min(X_1, \dots, X_n)$
 $f_{Y_1}(y) = n f(y) [S(y)]^{n-1}$
 $S_{Y_1}(y) = \prod_{i=1}^n \Pr(X_i > y)$
- $Y_n = \max(X_1, \dots, X_n)$
 $f_{Y_n}(y) = n f(y) [F(y)]^{n-1}$
 $F_{Y_n}(y) = \prod_{i=1}^n \Pr(X_i \leq y)$
- $Y_k \in (Y_1, \dots, Y_k, \dots, Y_n)$
 $f_{Y_k}(y) = \frac{n! \cdot f(y) [F(y)]^{k-1} [S(y)]^{n-k}}{(k-1)!(n-k)!}$
 $F_{Y_k}(y) = \Pr(\text{at least } k \text{ of } n X_i \text{ are } \leq y)$
 $= \sum_{i=k}^n \binom{n}{i} [F(y)]^i [S(y)]^{n-i}$
- $x + y = \min(x, y) + \max(x, y)$, since one is for sure the max and the other the min.

Mode : Most likely probability

- $g(x) = f(x)$ or some time $g(x) = \ln f(x)$
- Mode** is the x that respects : $g'(x) = 0$

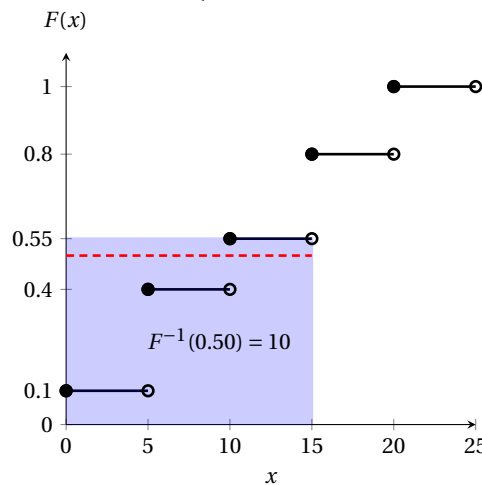
Normal Approximation

- $F_X(x) = \Phi\left(\frac{X - E[X]}{\sqrt{\text{Var}(X)}}\right)$
- Continuity correction** is necessary when X is discrete. $F_X(x) = \Phi\left(\frac{(X+k) - E[X]}{\sqrt{\text{Var}(X)}}\right)$ where k is the mid-point of the discrete value.

Discrete Cumulative Function

$$\Pr(X = x) = \begin{cases} 0.10, & x = 0 \\ 0.30, & x = 5 \\ 0.15, & x = 10 \\ 0.25, & x = 15 \\ 0.20, & x = 20 \end{cases}$$

$$\Pr(X \leq x) = \begin{cases} 0.10, & 0 \leq x < 5 \\ 0.40, & 5 \leq x < 10 \\ 0.55, & 10 \leq x < 15 \\ 0.80, & 15 \leq x < 20 \\ 1, & x \geq 20 \end{cases}$$



Contract

- Deductible(d)**
- Maximum(u)**
- Inflation(r)**
- Coinurance(α)**

$$Y = \begin{cases} 0 & x \leq \frac{d}{1+r} \\ \alpha[(1+r)x - d] & \frac{d}{1+r} < x < \frac{u}{1+r} \\ \alpha[u - d] & x \geq \frac{u}{1+r} \end{cases}$$

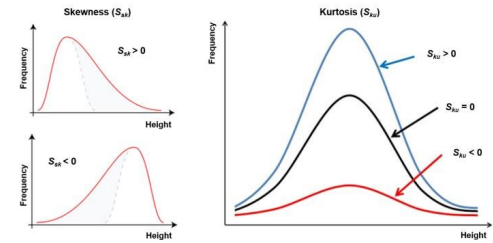
Warning : The maximal don't include the deductible.

Moments

- k^e moment about the origin. $\mu'_k = E[X^k]$
- k^e moment about the mean. $\mu_k = E[(X - \mu)^k]$

- The **Skewness** moment give information about the asymmetry of the distribution. If $S_{sk} = 0$, the distribution is normal.

$$S_{sk} = E\left[\left(\frac{X - \mu}{\sigma^2}\right)^3\right]$$



- The **kurtosis** moment give information about the flattening of the distribution. If $S_{ku} = 0$, the distribution is normal.

$$S_{ku} = E\left[\left(\frac{X - \mu}{\sigma^2}\right)^4\right]$$

- The **coefficient of variation** give information about the dispersion of the distribution.

$$CV = \frac{\sigma}{E[X]}$$

Transformations of distribution

- Lognormal** : $Y = e^X$, where
 $Y \sim \text{Lognormal}(\mu, \sigma)$
 $X \sim \text{Normal}(\mu, \sigma)$
- Inverse Exponential** : $Y = \frac{1}{X}$, where
 $Y \sim \text{Inverse Exponential}(1/\theta)$
 $X \sim \text{Exponential}(\theta)$
- Weibull** : $Y = X^{1/\tau}$, where
 $Y \sim \text{Weibull}(\sqrt[\tau]{\theta})$
 $X \sim \text{Exponential}(\theta)$

Parameter interpretation

- Scale parameter** (θ, β, σ) : Affect the spread of the distribution.
- Rate parameter** (λ) : Affect the rate of data at mean. (1/scale)
- Shape parameter** (α, τ, γ) : Affect the shape rather than simply shift the distribution.

Produit de convolution

The convolution of 2 random variable is define as the sum of the two.

$$f_{X_1+X_2}(x) = \int_{-\infty}^{\infty} f_{X_1}(x-s) f_{X_2}(s) ds$$

$$F_{X_1+X_2}(x) = \int_{-\infty}^x F_{X_1}(x-s) f_{X_2}(s) ds$$