UNS
Master MathMods/MPA

Année 2014-2015
Yannick Baraud

Mathematical Statistics

3 hours

1. Give the Central Limit Theorem (assumptions and result).
   Donner le théorème de la limite centrale (hypothèses et conclusions)

2. Give the law of large numbers (assumptions and result)
   Donner la loi des grands nombres (hypothèses et conclusion).

3. Give Cochran's Theorem.
   Donner le théorème de Cochran

Exercise 1 Let $X_1, \ldots, X_n$ be i.i.d. random variables with uniform distribution on $[0, \theta]$ with $\theta > 0$.
Soit $X_1, \ldots, X_n$ *i.i.d. de loi uniforme sur* $[0, \theta]$ *avec* $\theta > 0$.

4. What is the maximum likelihood estimator $\widehat{\theta}_n$ of θ?
   Quel est l'estimateur du maximum de vraisemblance $\widehat{\theta}_n$ *de* θ*?*

5. Compute the cumulative distribution function of $\widehat{\theta}_n$.
   Calculer la fonction de répartition de $\widehat{\theta}_n$

6. Compute the density of $\widehat{\theta}_n$.
   Calculer la densité de $\widehat{\theta}_n$

7. For which value of $a \in \mathbb{R}$, $a\widehat{\theta}_n$ is unbiased.
   Pour quelle valeur de $a \in \mathbb{R}$, $a\widehat{\theta}_n$ *est-il sans biais?*

8. For which value of $b \in \mathbb{R}$, the quadratic risk of $b\widehat{\theta}_n$ is minimum?
   Pour quelle valeur de $b \in \mathbb{R}$, *le risque quadratique de* $b\widehat{\theta}_n$ *est-il minimum?*

Exercise 2 Let $X_1, \ldots, X_n$ be i.i.d. Gaussian random variables with mean $\theta \in \mathbb{R} = \Theta$ and variance 1.
soit $X_1, \ldots, X_n$ *i.i.d. gaussiènne de moyenne* $\theta \in \mathbb{R} = \Theta$ *et de variance 1.*

9. Give an estimator $\widehat{\theta}_n$ of θ based on the moment method.
   Donner un estimator $\widehat{\theta}_n$ *de* θ *basé sur la méthode des moments.*

10. Give the estimator $\widetilde{\theta}_n$ of θ based on the median.
    Donner un estimator $\widetilde{\theta}_n$ *de* θ *basé sur la médiane.*

11. Show that $\sqrt{n}\left(\widetilde{\theta}_n - \theta\right)$ converges to some limit distribution and specify the limit.
Montrer que $\sqrt{n}\left(\widetilde{\theta}_n - \theta\right)$ *converge en loi et préciser la limite.*

We now use a Bayesian approach to estimate θ and use the prior ν_τ on Θ given by the Gaussian distribution with mean 0 and variance $\tau^2 > 0$.
On utilise à présent une approche bayésienne pour estimer θ basée sur la loi a priori ν_τ qui est gaussiènne de moyenne 0 et de variance $\tau^2 > 0$

12. Show that the posterior distribution is Gaussian with mean $\frac{n\tau^2}{n\tau^2+1}\overline{X}_n$ and variance $\frac{\tau^2}{n\tau^2+1}$
Montrer que la loi a posteriori est gaussiènne de moyenne $\frac{n\tau^2}{n\tau^2+1}\overline{X}_n$ et de variance $\frac{\tau^2}{n\tau^2+1}$

13. Compute the Bayes estimator $\widehat{\theta}_n^\tau$ associated to the prior ν_τ.
Calculer l'estimateur bayésien $\widehat{\theta}_n^\tau$ associé à la loi a priori ν_τ

14. Compute the quadratic risk $R(\theta, \overline{X}_n)$ of $\overline{X}_n$ at θ.
Calculer le risque quadratique $R(\theta, \overline{X}_n)$ de $\overline{X}_n$ en θ

15. Compute the quadratic risk $R(\theta, \widehat{\theta}_n^\tau)$ of $\widehat{\theta}_n^\tau$ at θ.
Calculer le risque quadratique $R(\theta, \widehat{\theta}_n^\tau)$ de $\widehat{\theta}_n^\tau$ en θ

Given an estimator $\overline{\theta}_n$ of θ, we denote by $R^\tau(\overline{\theta}_n)$, the Bayes risk of $\overline{\theta}_n$ with respect to the prior ν_τ.
Etant donné un estimateur $\overline{\theta}_n$ de θ, on note $R^\tau(\overline{\theta}_n)$, the risk de Bayes de $\overline{\theta}_n$ associé à la loi a priori ν_τ

16. Compute $R_n = R^\tau(\overline{X}_n)$ and show that this quantity is independent of τ.
Calculer $R_n = R^\tau(\overline{X}_n)$ et montrer que cette quantité est indépendante de τ.

17. Compute $R^\tau(\widehat{\theta}_n^\tau)$
Calculer $R^\tau(\widehat{\theta}_n^\tau)$

18. Show that
Montrer que
$$\lim_{\tau \to +\infty} R^\tau(\widehat{\theta}_n^\tau) = \sup_{\theta \in \Theta} R(\theta, \overline{X}_n)$$

19. Show that $\overline{X}_n$ is minimax.
Montrer que $\overline{X}_n$ est minimax.

Exercise 3 Let $X_1, \ldots, X_n$ be i.i.d. random variables with density
Soit $X_1, \ldots, X_n$ *i.i.d. de densité*

$$f_\theta(x) = \frac{\theta}{x^{\theta+1}} \mathbb{1}_{x \geq 1} \quad \text{with} \quad \theta \in \Theta = (0, +\infty).$$

20. Compute the cumulative distribution function of X_1.
    Calculer la fonction de répartition de X_1

21. Show that the statistical model is an exponential family.
    Montrer que le modèle statistique est une famille exponentielle

22. What is the maximum likelihood $\widehat{\theta}_n$ estimator of θ?
    Quel est l'estimateur du maximum de vraisemblance?

23. What is the limit in distribution of $\sqrt{n}\left(\widehat{\theta}_n - \theta\right)$?
    Quelle est la limite en loi de $\sqrt{n}\left(\widehat{\theta}_n - \theta\right)$ *?*

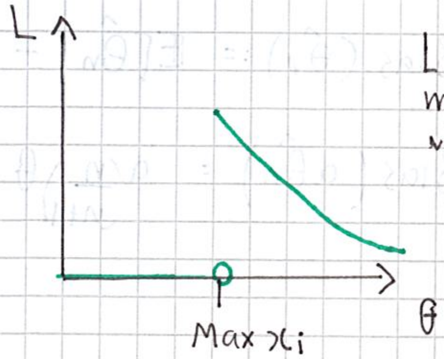
24. Build an asymptotic two-sided confidence interval for θ.
    Construire un intervalle de confiance asymptotique bilatère pour θ

Exercise 1 $X_1, \dots, X_n$ iid $U[0, \theta]$

$L = f(x_1) f(x_2) \cdots f(x_n)$

if $\max x_i > \theta$ $L = 0$

if $\max x_i \le \theta$ $L = \left(\frac{1}{\theta}\right)^n$



L is maximum when max x_i

Max Likelihood Estimator $\hat{\theta}_n = \max x_i$

Comulative distribution function of $\hat{\theta}_n = \max x_i$

$$F_{\hat{\theta}_n}(x) = \mathbb{P}(X_1 \le x, X_2 \le x, \dots X_n \le x)$$

$$\overset{ind}{=} \mathbb{P}(X_1 \le x)\, \mathbb{P}(X_2 \le x) \cdots \mathbb{P}(X_n \le x)$$

$$\mathbb{P}(X < x) = \int_{-\infty}^{x} \frac{1}{\theta} \mathbb{1}_{(0,\theta)} dx$$

if $0 \le x \le \theta$ $= \frac{x}{\theta}$

if $x < 0$ $= 0$

if $x > \theta$ $= 1$

if $x < 0$ $F(x) = 0$

if $0 \le x \le \theta$ $F(x) = \left(\frac{x}{\theta}\right)^n$

if $x > \theta$ $F(x) = 1$

$f(x) = \frac{dF(x)}{dx} =$

if $x < 0$ $f(x) = 0$

if $0 \le x \le \theta$ $f(x) = n\left(\frac{x^{n-1}}{\theta^n}\right)$

if $x > \theta$ $f(x) = 0$

$$E[\hat{\theta}_n] = \int_{-\infty}^{\infty} x f(x)\, dx = \int_0^{\theta} x\, n\left(\frac{x^{n-1}}{\theta^n}\right) dx = \frac{n\, x^{n+1}}{(n+1)\theta^n}\Bigg|_0^{\theta}$$

$$= \frac{n\, \theta^{n+1}}{(n+1)\theta^n} = \frac{n}{n+1}\theta \Rightarrow \mathbb{E}[a\hat{\theta}_n] = \frac{an}{(n+1)}\theta$$

$$\text{Bias}(\hat{\theta}_n) := \mathbb{E}[\hat{\theta}_n] - \theta$$

- $\text{Bias}(a\hat{\theta}_n) = a\left(\frac{n}{n+1}\right)\theta - \theta \Rightarrow$ unbiased if $a\frac{n}{n+1}\theta - \theta = 0$

$$a = \left(\frac{n+1}{n}\right)$$

- $\text{Risk}(\hat{\theta}_n) = \text{Bias}^2(\theta_n) + \text{var}(\hat{\theta}_n)$

$$= \cancel{\left(\left(\frac{n}{n+1}\right)\theta\right)^2 + \frac{n\theta^2}{n+2}}$$

$$= \left[\left(\frac{n}{n+1}\right)\theta - \theta\right]^2 + \frac{n\theta^2}{n+2} - \left(\frac{n\theta}{n+1}\right)^2$$

$\text{Risk}(b\hat{\theta}_n) =$

$$\left[b\left(\frac{n}{n+1}\right)\theta - \theta\right]^2 + \left(\frac{b^2 n\theta^2}{n+2} - \frac{b^2 n^2 \theta^2}{(n+1)^2}\right)$$

$$\text{var}(\hat{\theta}_n) = E[\theta_n^2] - (E[\theta_n])^2$$

$$E[\theta_n^2] = \int_{-\infty}^{\infty} x^2 f(x)\,dx$$

$$= \int_0^{\theta} x^2 \frac{n x^{n-1}}{\theta^n}\,dx$$

$$= n \frac{x^{n+2}}{(n+2)\theta^n}\Bigg|_0^{\theta}$$

$$= \frac{n\theta^2}{n+2}$$

$$\text{var}(\hat{\theta}_n) = \frac{n\theta^2}{n+2} - \left(\frac{n}{n+1}\theta\right)^2$$

$$\mathbb{E}[b\hat{\theta}_n]^2 = b^2\left(\frac{n}{n+2}\right)\theta^2$$

$$\text{var}(b\hat{\theta}_n) = b^2\left(\frac{n}{n+2}\right)\theta^2 - \left(\frac{bn}{n+1}\theta\right)^2$$

$$= b^2\frac{n\theta^2}{(n+2)} - \frac{b^2 n^2 \theta^2}{(n+1)^2}$$

Exercise 2

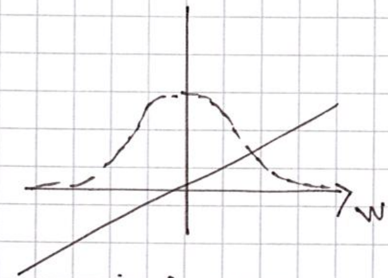
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}}$$

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}} dx$$

$w = x - \theta$

$dw = dx$

$$= \int_{-\infty}^{\infty} \frac{(w+\theta)}{\sqrt{2\pi}} e^{-\frac{(w)^2}{2}} dx$$



pair function

times impair function

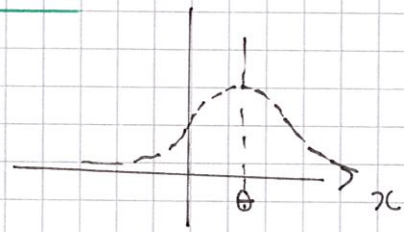
$$= \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} w e^{-w^2/2} dx}_{=0} + \underbrace{\frac{\theta}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-w^2/2} dx}_{=\theta}$$

$$E[X] = \theta$$

by Moment Method $\Rightarrow$ $\hat{\theta}_n = \bar{X}_n$

median x s.t $F(x) = \frac{1}{2}$

$q\left(\frac{1}{2}\right) = m = \theta$



symmetric on θ

$$\hat{\theta}_n = \hat{q}\left(\tfrac{1}{2}\right)$$

Thm $\sqrt{n}\left(\hat{q}_n(u) - q(u)\right) \xrightarrow{L} N\left(0, \frac{u(1-u)}{\left(F'(q(u))\right)^2}\right)$

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{L} N\left(0, \frac{\frac{1}{4}}{\left(\frac{1}{\sqrt{2\pi}}\right)^2}\right)$$

$F'(q(u)) = f(q(u))$

$= \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-q(u))^2}{2}}$

$q(\frac{1}{2}) = 0 \Rightarrow f(q(\frac{1}{2})) = \frac{1}{\sqrt{2\pi}}$

$$\xrightarrow{L} N\left(0, \frac{\pi}{2}\right)$$

Posterior Gaussian

$$f(x|\mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}} \qquad \pi(\mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\mu^2}{2\tau^2}}$$

$$\pi(\mu)\, L_n f(x^n|\mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\mu^2}{2\tau^2}} \frac{1}{(2\pi)^{n/2}} e^{\sum_i^n \frac{(x_i-\mu)^2}{2}}$$

$$\propto e^{-\frac{\mu^2}{2\tau^2} - \frac{1}{2}\sum_{i=1}^n (x_i^2 + \mu^2 - 2x_i\mu)} \qquad \bar{x} = \frac{1}{n}\sum x_i$$

$$\propto e^{-\frac{1}{2}\left[\frac{\mu^2}{\tau^2} + n\mu^2 - 2\mu\bar{x}n + \sum_i^n x_i^2\right]}$$

$$\propto e^{-\frac{1}{2}\left(\frac{1}{\tau^2}+n\right)\left[\mu^2 - \frac{2\mu n\bar{x}}{\left(\frac{1}{\tau^2}+n\right)} + \frac{\sum^n x_i^2}{\left(\frac{1}{\tau^2}+n\right)}\right]}$$

$$\frac{1}{\tau^2} + n = \frac{1 + n\tau^2}{\tau^2}$$

$$\propto e^{-\frac{1}{2\left(\frac{\tau^2}{1+n\tau^2}\right)}\left[\mu^2 - 2\mu n\bar{x}\left(\frac{\tau^2}{n\tau^2+1}\right) + \sum x_i^2\left(\frac{\tau^2}{1+n\tau^2}\right)\right]}$$

variace $\nu^2 = \left(\frac{\tau^2}{1+n\tau^2}\right)$

$$\propto e^{-\frac{1}{2\nu^2}\left[-2\mu n\bar{x}\nu^2 + \mu^2 + \sum^n x_i^2 \nu^2\right]}$$

$$\propto e^{-\frac{1}{2\nu^2}\left[\mu^2 - 2\mu n\bar{x}\nu^2 \pm (n\bar{x}\nu^2)^2\right] - \frac{1}{2}\sum x_i^2}$$

$$\propto e^{-\frac{1}{2\nu^2}\left[\mu - n\bar{x}\nu^2\right]^2 + \underbrace{\frac{1}{2}n^2\bar{x}^2\nu^2 - \frac{1}{2}\sum x_i^2}_{\text{constant}}}$$

μ_p mean $= n\bar{x}\nu^2$

$= \bar{x}\frac{n\tau^2}{1+n\tau^2}$

$$\Rightarrow f(\mu|x^n) \overset{d}{\sim} N(\mu_p, \nu^2)$$

under l (square loss function)
Bayes Estimator $\hat{\mu}$ is equal is the posteriori mean
because minimizing Bayes risk is equivalent
to minimize the posteriori mean

$$\frac{dR}{d\hat{\theta}} = \frac{d}{d\hat{\theta}}\int (\theta-\hat{\theta})^2 f(\theta|\mu)\, d\theta = 2\int (\theta-\hat{\theta}) f(\theta|\mu)\, d\theta$$

$$= 2\int \theta f(\theta|\mu)\, d\theta - 2\hat{\theta}\underbrace{\int f(\theta|\mu)\, d\theta}_{1} = 0$$

$$\Rightarrow \hat{\theta} = \int \theta f(\theta|\mu)\, d\theta = \mathbb{E}_p[\theta]$$

Bayes Est. $\hat{\mu}_p = \left(\frac{n\tau^2}{n\tau^2+1}\right)\bar{x}$

Quadratic Risk $\bar{X}_n$
Quadratic Risk $\hat{\mu}_p$

Bias $\bar{X}_n$: $\mathbb{E}\bar{X}_n - \mu = 0$
var $\bar{X}_n$: $\frac{1}{n^2} n \, var(X) = \frac{1}{n}$
$\Big\}\ R(\bar{X}_n, \mu) = \frac{1}{n}$

$$\text{Bias } \hat{\mu}_p = E[\hat{\mu}_p] - \mu = \frac{n\tau^2}{n\tau^2+1} E\bar{X}_n - \mu = \frac{n\tau^2\mu - \mu(n\tau^2+1)}{n\tau^2+1}$$

$$= -\frac{\mu}{n\tau^2+1}$$

$$var(\hat{\mu}_p) = \left(\frac{n\tau^2}{n\tau^2+1}\right)^2 var\,\bar{X}_n = \left(\frac{n\tau^2}{n\tau^2+1}\right)^2 \frac{1}{n} = \frac{n\tau^4}{(n\tau^2+1)^2}$$

$$\text{Quadratic Risk } \hat{\mu}_p : \frac{\mu}{(n\tau^2+1)^2} + \frac{n\tau^4}{(n\tau^2+1)^2} = \frac{\mu + n\tau^4}{(n\tau^2+1)^2}$$

Bayes risk r

$$r(\pi(\theta), \hat{\theta}) = \int R(\theta, \hat{\theta}) \pi(\theta) d\theta$$

$$r(\pi(\mu), \bar{X}_n) = \int \frac{1}{n} \frac{1}{\sqrt{2\pi}} e^{-\mu^2/2\tau^2} d\mu = \frac{1}{n}$$

$$\underline{\underline{r(\pi(\mu), \bar{X}_n) = \frac{1}{n}}}$$

$$r(\hat{\mu}_p, \pi(\mu)) = \int \frac{\mu + n\tau^4}{(n\tau^2+1)^2} \frac{1}{\sqrt{2\pi}} e^{-\mu^2/2\tau^2} d\mu$$

$$\underline{\underline{r(\pi(\mu), \hat{\mu}_p) = \frac{n\tau^4}{(n\tau^2+1)^2}}}$$

$$\sup_{\theta} R(\theta, \bar{X}_n) = \frac{1}{n}$$

$$\inf_{\hat{\theta}} \sup_{\theta} R(\theta, \bar{X}_n) = \frac{1}{n}$$

$$\lim_{\tau \to \infty} R(\hat{\mu}_p) = \lim_{\tau \to \infty} \frac{n\tau^4}{(n\tau^2+1)^2}$$

$$= \lim_{\tau \to \infty} \frac{4n\tau^3}{4n^2\tau^3 + 4n\tau} =$$

$$= \lim_{\tau \to \infty} \frac{1}{n} = \frac{1}{n}$$

Exercise 3

$$f(x) = \frac{\theta}{x^{\theta+1}} \mathbb{1}_{x \geq 1} \qquad \theta \in (0, \infty)$$

Cumulative

$$F(x) = \int_{-\infty}^{\infty} f(x)\,dx = \int_1^{\infty} \frac{\theta}{x^{\theta+1}}\,dx = \int_1^{\infty} \theta x^{-(\theta+1)}\,dx$$

$$= \frac{\theta x^{1-(\theta+1)}}{1-(\theta+1)}\Bigg|_1^{\infty} = \frac{\cancel{\theta} x^{-\theta}}{-\cancel{\theta}}\Bigg|_1^{\infty} = -\frac{1}{x^{\theta}}\Bigg|_1^{\infty} = 1$$

Exponential family

$$\frac{\theta}{x^{\theta+1}} = \exp\left[\log\left(\frac{\theta}{x^{\theta+1}}\right)\right] = \exp\left[\log\theta - \log x^{\theta+1}\right]$$

$$= \exp\left[\log\theta - (\theta+1)\log x\right]$$

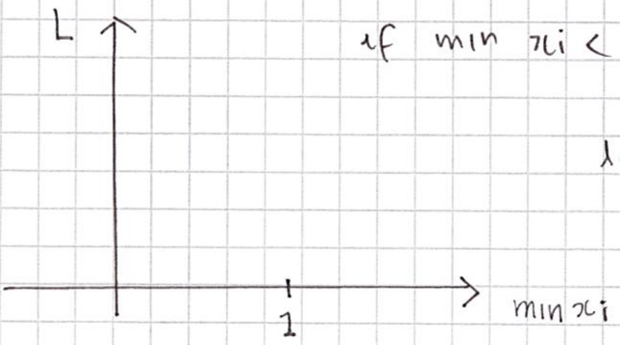
$$= \theta\, e^{-\theta \log x} e^{-\log x} = \theta\, e^{-\theta\log x} \frac{1}{x}$$

$$= \alpha(\theta)\, e^{\beta(\theta)T(x)}\, h(x)$$

Maximum Likelihood

$$L = f(x_1) f(x_2) \cdots f(x_n) =$$

if $\min x_i \geq 1$ $\quad L = \dfrac{\theta^n}{(x^{\theta+1})^n}$



if $\min x_i < 1$ $\quad L = 0$

$$\log L = n\log\theta - n(\theta+1)\log x$$

$$\frac{d \log L}{d\theta} = n\frac{1}{\theta} - n\log x = 0$$

$$\Rightarrow \hat{\theta}_n = (\log x)^{-1}$$